

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису  
УДК 004.942:519.216.3

«До захисту допущено»  
В. о. завідувача кафедри ММСА  
\_\_\_\_\_ О. Л. Тимошук  
“    ” \_\_\_\_\_ 2020 р.

## Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз  
на тему: «Моделювання та оцінювання операційного фінансового ризику»

Виконала:

студентка II курсу, групи КА-91мп

Корнійчук Анна Сергіївна \_\_\_\_\_

Керівник:

професор кафедри ММСА,

д.т.н., проф. Бідюк П. І. \_\_\_\_\_

Рецензент:

декан ФІОТ КПІ ім. Ігоря Сікорського,

професор, д.т.н. Теленик С. Ф. \_\_\_\_\_

Засвідчую, що у цій магістерській  
дисертації немає запозичень з праць інших  
авторів без відповідних посилань.

Студент \_\_\_\_\_

Київ  
2020

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
 «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»  
 ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
 КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)  
 Спеціальність — 124 «Системний аналіз»

ЗАТВЕРДЖУЮ  
 В. о. завідувача кафедри ММСА  
 О. Л. Тимощук  
 «\_\_\_» \_\_\_\_\_ 2020 р.

### ЗАВДАННЯ

на магістерську дисертацію студенту Корнійчук Анні Сергіївні

**1. Тема дисертації:** «Моделювання та оцінювання операційного фінансового ризику», науковий керівник дисертації Бідюк Петро Іванович, д.т.н., професор, затверджені наказом по університету від «02» листопада 2020 р. № 3182-с.

**2. Термін подання студентом дисертації:** 16 грудня 2020 р.

**3. Об'єкт дослідження:** нелінійні нестационарні процеси у фінансах, пов'язані з операційними ризиками, які представлені відповідними статистичними даними та/або експертними оцінками стосовно особливостей їх розвитку, операційний ризик.

**4. Предмет дослідження:** ймовірно-статистичні та методи інтелектуального аналізу даних для моделювання і оцінювання ризиків можливих операційних втрат; множини критеріїв для аналізу адекватності моделей.

**5. Перелік завдань, які потрібно розробити:**

- 1) дослідити актуальність проблеми моделювання і оцінювання операційних ризиків;
- 2) зробити огляд літератури, результатів відомих досліджень на обрану тему;
- 3) виконати огляд методів моделювання та оцінювання операційних ризиків;
- 4) зібрати необхідні статистичні дані для виконання обчислювальних експериментів;
- 5) виконати моделювання та оцінювання операційного ризику за допомогою мережі Байєса та дерев рішень;
- 6) виконати аналіз отриманих результатів, сформулювати перспективи подальших досліджень;

- 7) розробити стартап-проект виведення на ринок результатів дослідження;
- 8) розробити концептуальні висновки за результатами наукового дослідження.

#### **6. Орієнтовний перелік графічного (ілюстративного) матеріалу:**

- 1) інтерфейси програмного забезпечення для роботи із операційними ризиками;
- 2) результати роботи створених моделей;
- 3) таблиці у розділі стартап-проекту.

#### **7. Орієнтовний перелік публікацій:**

Участь у Восьмій міжнародній науково-технічній конференції з публікацією тез доповіді. Стаття у збірнику ІПСА «Системні науки і кібернетика».

**8. Дата видачі завдання:** 02 вересня 2020 р.

### **Календарний план**

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів. Дослідження актуальності даної проблеми.	05.09.2020—10.09.2020
2.	Перший розділ. Огляд літератури. Понятійно-категоріальний апарат. Огляд існуючих методів моделювання та оцінювання ОР.	11.09.2020—25.09.2020
3.	Другий розділ. Теоретичні основи байєсівських мереж, методи їх побудови. Огляд програмних середовищ для побудови БМ. Теорія побудови дерев рішень.	26.09.2020—15.10.2020
4.	Третій розділ. Пошук та обробка даних. Побудова мережі Байєса. Побудова дерев рішень. Аналіз результатів та підбір ілюстративних прикладів.	16.10.2020—07.11.2020
5.	Четвертий розділ. Розробка стартап-проекту.	08.11.2020—11.11.2020
6.	Концептуальні висновки. Перспективи розвитку отриманих рішень.	12.11.2020—16.11.2020
7.	Оформлення дисертації та підготовка ілюстративного матеріалу для доповіді.	17.11.2020—29.11.2020

Студент

А.С. Корнійчук

Науковий керівник дисертації

П.І. Бідюк

## РЕФЕРАТ

Магістерська дисертація: 109 с., 21 табл., 22 рис., 3 дод., 36 джерел.

ФІНАНСОВІ ВТРАТИ, ОПЕРАЦІЙНІ ФІНАНСОВІ ВТРАТИ, ОПЕРАЦІЙНИЙ РИЗИК, БАЗЕЛЬ ІІ, БАЙЄСІВСЬКІ МЕРЕЖІ, ТЕОРЕМА БАЙЄСА, ІМОВІРНІСНИЙ ВИСНОВОК, АЛГОРИТМ HUGIN, ДЕРЕВА РІШЕНЬ, ВИПАДКОВИЙ ЛІС, CART.

Об'єктом дослідження є нелінійні нестационарні процеси у фінансах, пов'язані з операційними ризиками, які представлені відповідними статистичними даними та/або експертними оцінками стосовно особливостей їх розвитку, операційний ризик.

Предметом дослідження є ймовірно-статистичні та методи інтелектуального аналізу даних для моделювання і оцінювання ризиків можливих операційних втрат; множини критеріїв для аналізу адекватності моделей.

Метою роботи є аналіз методів моделювання та оцінювання операційного ризику та побудова моделей для оцінювання операційних ризиків, перевірка їх ефективності.

Методи дослідження: мережі Байєса, алгоритм Hugin формування ймовірного висновку, дерева рішень CART і випадковий ліс.

У роботі наведено огляд існуючих методів моделювання та оцінювання операційних ризиків, а також деякі сучасні програмні продукти для роботи з ними. Детально розглянуто мережі Байєса як метод моделювання та оцінювання операційних ризиків, а також дерева рішень. Застосування цих методів проілюстровано відповідними результатами.

Науковою новизною роботи є створена на основі експертних даних мережа Байєса та навчені на отриманих з її допомогою даних дерева рішень. Модель на основі мережі Байєса може використовуватися для оцінювання необхідного капіталу на покриття операційних ризиків у комерційних банках.

## ABSTRACT

Master thesis: 109 p., 21 tabl., 22 fig., 3 appendixes, 36 sources.

FINANCIAL LOSSES, OPERATIONAL FINANCIAL LOSSES, OPERATIONAL RISK, BASEL II, BAYESIAN NETWORKS, BAYES' THEOREM, INFERENCE, HUGIN ALGORITHM, DECISION TREES, RANDOM FOREST, CART.

The object of research are non-linear non-stationary processes in finance, related to financial risks, which are presented by relevant statistics and/or expert assessments regarding the specific features of their development, operational risk.

The subject of research is probabilistic-statistical and data mining methods for modeling and assessing the risks of possible operational losses; sets of criteria for analyzing the adequacy of models.

The purpose of the work is to consider methods of modeling and assessment of operational risk and construction of models for assessing operational risks, checking their effectiveness.

Research methods: Bayesian networks, Hugin inference algorithm, CART decision tree and random forest.

The paper provides an overview of existing methods of modeling and assessment of operational risks, as well as some modern software products for working with them. Bayesian networks as a method of modeling and assessing operational risks, as well as decision trees, are considered in detail. The application of these methods is illustrated by the corresponding results.

The scientific novelty of the work is the Bayesian network created on the basis of expert data and decisions made on the basis of data obtained with its help. The model on the basis of Bayesian network can be used to estimate the capital required to cover operational risks in commercial banks.

## ЗМІСТ

<b>ПЕРЕЛІК ПРИЙНЯТИХ СКОРОЧЕНЬ</b>	9
<b>ВСТУП</b>	10
<b>РОЗДІЛ 1 АНАЛІЗ АКТУАЛЬНОСТІ ПРОБЛЕМИ</b>	12
1.1 Поняття фінансового та операційного ризику	12
1.2 Аналіз актуальності проблеми моделювання і оцінювання ризиків операційних фінансових втрат	14
1.2.1 Приклади реалізації операційного ризику	14
1.2.2 Актуальність проблеми	15
1.2.3 Вплив пандемії на операційні ризики фінансових організацій	16
1.3 Огляд результатів відомих досліджень	17
1.4 Існуючі методи моделювання і оцінювання ризиків операційних фінансових втрат	20
1.5 Деякі сучасні системи для роботи з операційними ризиками	22
1.6 Висновки до розділу 1	31
<b>РОЗДІЛ 2 ВИБРАНІ МЕТОДИ І ПІДХОДИ ДО МОДЕЛЮВАННЯ І ОЦІНЮВАННЯ РИЗИКІВ ОПЕРАЦІЙНИХ ВТРАТ</b>	33
2.1 Ймовірнісний підхід - байесівські мережі	33
2.1.1 Теорема Байєса та обчислення ймовірностей	34
2.1.2 Означення мережі Байєса та класифікація	36
2.1.3 Методи побудови і оцінювання структури мереж Байєса	38
2.1.3.1 Методи на основі оціночних функцій	39
2.1.3.2 Методи на основі використання тестів на умовну незалежність	41
2.1.3.3 Інші методи	42

2.1.4 Формування імовірнісного висновку за алгоритмом Hugin	45
2.2 Древа рішень як альтернативний метод моделювання та оцінювання ризиків операційних втрат	49
2.2.1 Поняття дерева рішень та їх класифікація	49
2.2.2 Способи розділення вузлів	51
2.2.3 Переваги і недоліки дерев рішень	53
2.2.4 Випадковий ліс	54
2.3 Висновки до розділу 2	56
<b>РОЗДІЛ 3 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ</b>	57
3.1 Обґрунтування вибору платформи	57
3.2 Вимоги до обладнання та інструменти для роботи з даними	58
3.3 Опис архітектури СППР і функціональної схеми	59
3.4 Побудова мережі Байєса для оцінювання операційних ризиків	60
3.4.1 Визначення структури мережі	60
3.4.2 Визначення апріорних імовірностей	64
3.4.3 Обчислення необхідного капіталу на покриття ОР та сценарний аналіз	65
3.5 Оцінювання та прогнозування операційних ризиків за допомогою дерев рішень	68
3.5.1 Підготовка даних та вибір метрик	68
3.5.2 Побудова дерева рішень та аналіз результатів	72
3.6 Висновки до розділу 3	78
<b>РОЗДІЛ 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ</b>	79
4.1 Опис ідеї стартап-проекту	79
4.2 Розробка бізнес-моделі стартапу	80

4.3 Аналіз ринкових можливостей та розробка маркетингової стратегії стартап-проекту	82
4.4 Розробка маркетингової програми стартап-проекту	86
4.5 Висновки до розділу 4	90
<b>ВИСНОВКИ</b>	91
<b>ПЕРЕЛІК ПОСИЛАНЬ</b>	93
<b>ДОДАТОК А ЛІСТИНГ ПРОГРАМИ</b>	97
<b>ДОДАТОК Б ТАБЛИЦІ АПРІОРНИХ ІМОВІРНОСТЕЙ</b>	101
<b>ДОДАТОК В НАУКОВІ ПУБЛІКАЦІЇ</b>	109



## ПЕРЕЛІК ПРИЙНЯТИХ СКОРОЧЕНЬ

БМ – байєсівська мережа

ОР – операційний ризик

САГ – спрямований ациклічний граф

УМВ – упорядкована множина вершин

## ВСТУП

Операційний ризик притаманний всім фінансовим установам. Операційний ризик існував давно, проте відносився він до категорії «інші ризики», а його концепція була інтерпретована лише після 1995 року, коли один з найстаріших банків Лондона, Barings Bank, зазнав краху через одного з торговців внаслідок несанкціонованих спекуляцій.

За розробку рекомендацій щодо моделювання, оцінювання і менеджменту операційних ризиків узявся Базельський комітет з питань банківського нагляду, що відобразилося, зокрема, у документі «International Convergence of Capital Measurement and Capital Standards: A Revised Framework» або Базель II.

Так, операційний ризик став одним із найважливіших для фінансових установ, адже втрати від його реалізації можуть бути катастрофічними.

З початком пандемії не залишилося країн, які б не мали проблем із економікою. Багато фінансових установ були не готові до наслідків запровадження карантину та переходу на дистанційний режим роботи. На тлі теперішніх подій, коли безліч компаній і підприємств змушені призупиняти свою діяльність або взагалі виявляються банкрутами, важливою необхідністю є прийняття обґрунтованих управлінських рішень на основі високоякісних моделей ризиків і прогнозів розвитку фінансових процесів.

У зв'язку з поширенням COVID-19 посилився вплив проблем інформаційної безпеки, ведення документації та шахрайства, пов'язаного з кіберпростором, на роботу установ. Тому моделювання і оцінювання операційних ризиків є актуальною проблемою.

У першому розділі було розглянуто поняття фінансового та операційного ризиків, їх класифікацію, наведено огляд літератури та результатів відомих досліджень на дану тему, деякі існуючі методи моделювання та оцінювання

операційного ризику. Також було проведено детальний аналіз сучасного програмного забезпечення, яке вважається найпопулярнішим та найзручнішим для роботи із операційним ризиком.

У другому розділі роботи подано теорему Байєса, означення мереж Байєса, їх класифікацію. Крім цього зроблено огляд методів побудови і оцінювання структури мереж Байєса. Також детально розглянуто алгоритм Hugin для формування імовірнісного висновку за допомогою мережі Байєса. Крім цього розглянуто дерева рішень, такі як CART та випадковий ліс.

У третьому розділі виконується моделювання та оцінювання операційних ризиків за допомогою мереж Байєса та дерев рішень. Детально розглядаються етапи побудови цих моделей, а також наводяться відповідні результати, проілюстровані необхідними рисунками і таблицями.

У четвертому розділі наводяться етапи створення стартап-проекту на основі побудованої байєсівської мережі.

## РОЗДІЛ 1 АНАЛІЗ АКТУАЛЬНОСТІ ПРОБЛЕМИ

У даному розділі розглянуто поняття фінансового та операційного ризиків, їх класифікацію, проведено аналіз актуальності моделювання та оцінювання операційних ризиків, що підкріплено деякими історичними даними про випадки реалізації операційного ризику, а також аналізом впливу пандемії на актуальність даної задачі. Крім цього наведено огляд результатів відомих досліджень, серед яких як загальні роботи, так і праці, присвячені конкретним методам оцінювання операційного ризику. Також зроблено огляд існуючих методів моделювання та оцінювання операційних ризиків. Наприкінці розділу подано огляд програмного забезпечення для роботи із операційними ризиками із наведенням його можливостей та переваг.

### 1.1 Поняття фінансового та операційного ризику

Фінансовий ризик — ризик, пов'язаний з імовірністю втрат будь-яких грошових коштів або їх недоодержанням [1]. Він притаманний організаціям та підприємствам майже в будь-якій сфері їх діяльності.

Серед фінансових ризиків виділяють [2]:

- кредитний — ризик непогашення кредиту, несплати процентів за ним;
- процентний — ризик, який виникає у разі зміни процентних ставок за кредитними ресурсами, що надаються;

- валютний — ризик можливий у разі зміни курсів валют, а також політичної ситуації, коли курси валют незмінні, а можливості вільного обігу валют обмежені. Ці ризики пов'язані з переоцінкою статей балансу закордонних філій підприємств у національну валюту та при зворотних операціях;

- галузевий — пов'язаний зі специфікою діяльності окремих галузей, проявляється в змінах інвестиційної якості та курсової вартості цінних паперів і відповідних втратах інвесторів;

- ліквідності та структури капіталу — ризик, пов'язаний з можливістю втрат при реалізації цінного папера через зміну в оцінці його якості;

- операційний — ризик пов'язаний із помилкою або неправильною організацією процесів в організації, неправильним вибором методу проведення тієї чи іншої фінансової операції;

- ризик країни — сукупність політичних, економічних та трансфертних ризиків і пов'язаний з дійсними та очікуваними політичними й економічними умовами в країні і впливом цих умов на здатність уряду країни, окремих корпорацій та фізичних осіб виконувати зобов'язання по зовнішньому боргу.

Отже, за означенням Базельського комітету, операційний ризик — це ризик прямих або непрямих втрат, викликаних помилками або недосконалістю процесів, систем в організації, помилками або недостатньою кваліфікацією персоналу організації або несприятливих зовнішніх подій нефінансової природи (наприклад, шахрайство або стихійне лихо). У це визначення входить і юридичний ризик (тобто ризик, який виникає через недотримання вимог законодавства, договорів, прийнятої практики, а також через можливість двозначного тлумачення законів та правил), але не входить стратегічний та репутаційний ризики [3].

Операційні ризики класифікують таким чином [4]:

- внутрішнє шахрайство – незаконне привласнення активів, ухилення від сплати податків, умисне неправильне зазначення посад, хабарництво;
- зовнішнє шахрайство – крадіжка інформації, злом, крадіжка і підробка третіми особами;
- практика найму і безпека на робочому місці – дискримінація, компенсація працівникам, здоров'я і безпека працівників;
- клієнти, продукти і ділова практика – маніпулювання ринком, антимонопольне законодавство, неналежна торгівля, дефекти продуктів, відтік рахунків;
- пошкодження фізичних активів – стихійні лиха, тероризм, вандалізм;
- збої в роботі і системні збої – збої в роботі комунальних служб, збої програмного забезпечення, устаткування;
- управління виконанням, доставкою та процесами – помилки введення даних, помилки бухгалтерського обліку, невиконання обов'язкової звітності, втрата клієнтських активів через необережність.

## 1.2 Аналіз актуальності проблеми моделювання і прогнозування ризиків операційних фінансових втрат

### 1.2.1 Приклади реалізації операційного ризику

Як відомо, операційний ризик притаманний всім фінансовим установам і потребує правильного менеджменту. Серед установ, які зазнали значних втрат через неправильний або не достатній менеджмент операційного ризику, є і відомі

фірми країн Європи та США. Зокрема розглянемо приклади реалізації операційного ризику, наведені у [5].

Одним з найвідоміших випадків несподіваних та величезних втрат, понесених установою у наслідок реалізації операційних ризиків, є ситуація з американським хедж-фондом Long-Term Capital Management (LTCM), банкрутство якого у 1998 році ледве не стало причиною колапсу світових фінансових ринків. Хоча на чолі організації стояв відомий інвестиційний банкір Джон Маріуезер, а його партнери Р. Мертон і М. Шоулз навіть здобули Нобелівську премію у галузі економіки, у результаті розробленої ними схеми інвестування на момент дефолту в Росії у 1998 році вони вели масштабні операції з державними цінними паперами і форвардними контрактами «рубль — долар». У наслідок цього фонд LTCM утратив 4 млрд доларів США при п'ятимільярдному капіталі фонду та активами під управління у 100 млрд доларів США. Банкрутство фонду стало б ударом для всієї світової фінансової системи, адже тоді LTCM мав відкриті позиції по всьому світу на загальну суму 1,25 трлн доларів США. Лише завдяки наданому консорціумом з 14 американських банків стабілізаційному кредиту на суму 3,6 млрд доларів США LTCM зміг устояти і, виплативши банкам борги, закритися у 1999 році.

Ще одним прикладом є ситуація з Allfirst (Allied Irish Banks), які зазнали збитків на 690 млн доларів США через шахрайські торгівельні схеми.

### 1.2.2 Актуальність проблеми

Операційні фінансові ризики є у будь-якій компанії і на підприємствах, притаманні фінансовим організаціям будь-яких типів, оскільки є інформаційні технології для документацій, помилки персоналу і т. ін.

Зокрема операційні ризики супроводжують діяльність комерційних банків. Вони також взаємопов'язані із іншими видами ризику, зокрема ринковим, кредитним. Проте, на відміну від цих ризиків, операційний ризик не виникає добровільно, не залежить від доходів і не може бути диверсифікований. Оскільки операційний ризик, зокрема, пов'язаний із людським фактором та недосконалістю систем, він не може бути повністю усунутим [6]. Зважаючи на численні випадки фінансових втрат та навіть банкрутства установ у наслідок реалізації операційного ризику, сучасні підприємства, фірми, банки зацікавлені у моделюванні, якісному оцінюванні та ефективному менеджменті операційних ризиків і вважають це одним із першочергових завдань.

### 1.2.3 Вплив пандемії на операційні ризики фінансових організацій

З початком пандемії та порушенням стабільності у світі, як наслідку запровадження карантину, увага до операційного ризику зросла.

За результатами форуму передових практик в області операційних ризиків [7], інформаційна безпека, ведення документації та шахрайство, пов'язане з кіберпростором, є трьома основними засобами управління фірм, ослабленими масовим переходом на роботу з дому, викликаним пандемією COVID-19.

Було зазначено, що у нинішніх умовах фінансові установи змушені переходити від звичайного режиму роботи до повного антикризового управління.

Багато фірм були готові до переходу на роботу з дому, але, можливо, не в тому масштабі, який був потрібний для їх заходів у відповідь на пандемію. Ця ситуація, безперечно, несе у собі зростання ризиків. Тому залишається лише своєчасно приймати правильні управлінські рішення для мінімізації ризиків,



зокрема і операційного. Так, учасники форму рекомендують звернути увагу на наступні пункти.

Інформаційна безпека. Існує політика роботи із конфіденційними друкованими документами у межах офісу. Винесення їх за межі офісу, робота з ними та їх утилізація у домашніх умовах потребує додаткової оцінки та розробки політики.

Ведення записів і доказів. Оскільки сьогодні світовий ринок є нестабільним, наразі важливим є активізація ведення документації та доказів взаємодії з клієнтами для зменшення наслідків майбутніх суперечок. Багато задач, пов'язаних із записом і підтвердженням взаємодії з клієнтом, ще не повністю автоматизовані. Тому фірми повинні оцінити, чи виконуються ці процеси із належною якістю вручну, щоб захистити себе.

Шахрайство і фішинг. Оскільки, працюючи вдома, співробітники можуть користуватися домашніми комп'ютерами без належного захисту, який мають корпоративні, фірми повинні підвищувати обізнаність про кібер-ризики і забезпечувати оновлення програмного забезпечення для захисту від шкідливих програм і застосовувати виправлення для віртуальної приватної мережі (VPN).

Як було підсумовано на формі [7], зараз у зв'язку із нестабільною ситуацією у світі надійність контролю може піти на спад, і своєчасне отримання незалежного висновку про операційний ризик має велике значення для зосередження уваги на правильних пріоритетних областях.

### 1.3 Огляд результатів відомих досліджень

У «International Convergence of Capital Measurement and Capital Standards: A Revised Framework» або Базель II подається означення операційного ризику,

запропоновані підходи до його моделювання та оцінювання, зазначені вимоги до розміру нормативного капіталу для покриття операційного ризику [8].

Документи, видані Базельським комітетом з питань банківського нагляду, є одними із основних документів, які регулюють банківську діяльність.

У своїй книзі [9] Джеймс Різон стверджує, що операційному ризику піддаються всі організації, використовує приклади з секторів фінансів, залізничного транспорту, цивільної авіації та атомної енергетики і приходить до висновку, що аварії є не тільки результатом людської схильності до помилок, але і підтримуються організаційними особливостями роботи організації, які, у разі неправильного підходу, не можуть убезпечити організацію від людських помилок, промахів і, в разі шахрайства, зловмисних дій. Людська помилка може розглядатися як початок катастрофічної події, але без прихованої слабкості всередині організації подія не досягне катастрофічних масштабів. З цього ми робимо висновок, що прогнозування операційного ризику нерозривно пов'язане з передовою практикою управління і що вимірювання операційного ризику може бути здійснено тільки в тому випадку, якщо ефективність конкретних організаційних процесів управління ризиками і контролю регулярно оцінюється і включається в моделювання. Це різко контрастує з думкою про те, що моделювання OpRisk включає виключно дослідження статистичних явищ, що властиво багатьом сучасним моделям АМА, заснованих на даних.

Більш того, через це ми виявляємо, що управління операційним ризиком в рамках бізнес-процесів банків занадто часто відривається від моделей, розроблених для кількісної оцінки нормативного капіталу [10].

У роботі [10] описується використання гібридних динамічних байєсівських мереж (HDBN) для моделювання операційного ризику, з яким стикаються фінансові установи, з точки зору економічного капіталу. У ній описується методологія моделювання фінансових втрат в результаті умисних або випадкових подій і визначається їх здатність обходити заходи контролю і, врешті решт,

приводити до все більш серйозних фінансових наслідків. Представлений підхід фокусується на моделюванні причин та наслідків збитків за допомогою моделі динамічної байєсівської мережі, що продемонстровано на прикладі з реальними даними.

У роботі [11] наведено приклад застосування АМА на реальних даних із дослідженням багатьох розподілів, таких як, розподіл Пуассона, емпіричний, експоненційний, розподіл Вейбулла та інші. Також розглянуто вплив екстремальних спостережень на результат – обчислення регуляторного капіталу.

У роботі [12] описується застосування байєсівських мереж та діаграм впливу для вимірювання та керування певними операційними ризиками, такими як ризик обробки транзакцій та людський ризик (transactions processing risk and human risk). Також приділяється увага бектестингу і сценарному аналізу.

Книга [13], написана провідними спеціалістами по роботі із операційними ризиками, є фундаментальною працею, у якій детально розглядається операційний ризик і його регулювання, аналіз та управління.

У дослідженні [14] подано визначення операційного ризику, а також його застосування у банківській та страховій діяльності. Детальними прикладами проілюстровані запропоновані у Базельських рекомендаціях підходи до розрахунку капіталу на покриття операційного ризику.

У роботі [15] автор досліджує результати використання розширених підходів до вимірювання (АМА) як методу оцінки витрат на покриття операційного ризику для банків та страхових компаній в рамках парадигм Базеля II та з урахуванням законодавчих актів США. АМА, розроблена в цій дисертації, використовує моделі актуарних збитків, доповнені теорією екстремальних значень, для визначення емпіричної функції розподілу ймовірностей загального нарахування капіталу з точки зору різних класів зв'язок.

#### 1.4 Існуючі методи моделювання і оцінювання ризиків операційних фінансових втрат

У Базелі II [8] розглядаються такі методи оцінювання операційного ризику у банківській діяльності:

- а) підхід базового індикатора (BIA, Basic Indicator Approach);
- б) стандартизований підхід (TSA, The Standardized Approach) і альтернативний стандартизований підхід (ASA);
- в) розширені підходи (AMA, Advanced Measurement Approach), що включають в себе такі підходи:
  - 1) підхід внутрішнього виміру (IMA, Internal Measurement Approach);
  - 2) підхід на основі розподілу втрат (LDA, Loss Distribution Approach);
  - 3) підхід на основі моделювання сценаріїв (SBA, Scenario-based approach);
  - 4) підхід скорингових карт або бально-ваговий підхід (SCA, Scorecard Approach).

У 2015 році Базельським комітетом було прийнято рішення про виключення АМА з Базеля II у майбутньому, оскільки даний метод не підтвердив очікувань щодо простоти його використання. Замінить АМА Standardised Measurement Approach (SMA). Основною відмінністю даного підходу є включення у модель нової компоненти збитків (Loss Component), яка враховує особливості операційного ризику конкретного банку за статистикою його внутрішніх збитків за останні не менше 5 років. У модель також був внесений ряд інших нововведень,

наприклад, поповнення складових ВІ, зміна деяких розрахункових формул, введення ряду обмежень і ін. [6].

У грудні 2017 року було видано програмний документ «Basel III: Finalising post-crisis reforms», у якому затверджено новий метод для розрахунку необхідного капіталу на покриття операційних ризиків. Впровадження даного підходу в практику планується з січня 2022 року [6].

Крім виконання рекомендацій Базельського комітету банки мають право розробляти власні моделі для оцінювання ризиків, проте вони потребують затвердження. На основі власних даних про реалізацію операційного ризику, а також за допомогою закритих баз даних, банки можуть розробляти та впроваджувати ефективні саме для їхньої діяльності методи роботи із операційними ризиками. На жаль, ці розробки здебільшого не є загальнодоступною інформацією.

Іншим підходом до роботи із операційними ризиками є застосування ймовірно-статистичних моделей у формі байєсівських мереж. Вони дають можливість особі, що приймає рішення, базувати свій вибір на «що буде, якщо...?» сценаріях для підкріплення рішення, є досить ефективними та зручними у використанні, що детально описано у літературі [12]. Обчислення капіталу, необхідного на покриття операційних ризиків, може проводитися за допомогою OpVaR (Operational Value-at-Risk), що є вираженою в грошових одиницях оцінкою величини, яку не перевищать можливі протягом певного періоду часу втрати із заданою ймовірністю. Рівень довіри, рекомендований Базельським комітетом, становить 99%. Однак для внутрішньої моделі управління ризиками це значення, зазвичай, становить 95%.

### 1.5 Деякі сучасні системи для роботи з операційними ризиками

Програмне забезпечення для управління операційними ризиками призначене для виявлення, оцінювання і усунення операційних ризиків у всіх відділах компанії. Цей тип програмного забезпечення використовується для запобігання втрат, які можуть бути викликані різними факторами, такими як поведінка людини, неузгоджені процеси або проблеми, пов'язані з інформаційними та бізнес-технологіями. Компанії використовують рішення по управлінню операційними ризиками для визначення і реалізації стратегій і планів, необхідних для оптимізації управління операційними ризиками. Працівники компаній використовують інструменти управління операційними ризиками для обміну інформацією та спільної роботи по зниженню та запобіганню ризиків [16].

За допомогою сайту [16], що демонструє неупереджені відгуки про задоволеність користувачів певним програмним забезпеченням у своїх рейтингах та звітах, було отримано список найпопулярнішого програмного забезпечення для управління операційним ризиком у 2020 році.

Отож серед 113 пропозицій найпопулярнішими та найзручнішими у використанні вважаються такі:

1. LogicGate Risk Cloud.
2. SAI360.
3. RSA Archer.

Розглянемо детальніше лідера рейтингу – LogicGate Risk Cloud [17].

LogicGate Risk Cloud <sup>TM</sup> – це комплексний набір додатків, що поєднують гнучкість і готові функціональні можливості для агрегування, управління і зменшення всього ландшафту ризиків всередині організації. Володіючи

кількісним розумінням рівня ризику для своєї організації, керівники підприємств можуть розширити свої програми управління ризиками і взяти на себе більш розумні ризики та можливості.

Серед функцій цього ПЗ є можливість визначенням операцій, необхідних для управління ризиками, за допомогою Agile GRC, автоматизація, яка дозволяє зменшити кількість та вплив людських помилок. ПЗ є зручним для впорядкування документації та звітності компанії (Рисунок 1), а також відстеження та пом'якшення ситуацій, пов'язаних із управлінням ризиками (Рисунок 2).

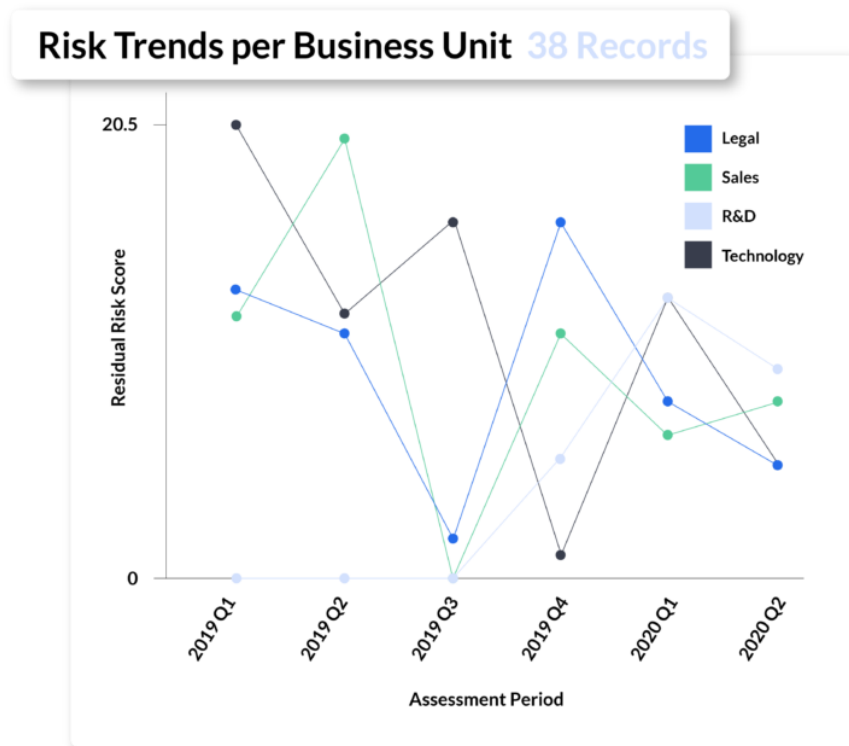


Рисунок 1 – Звітність щодо управління ризиками підприємств

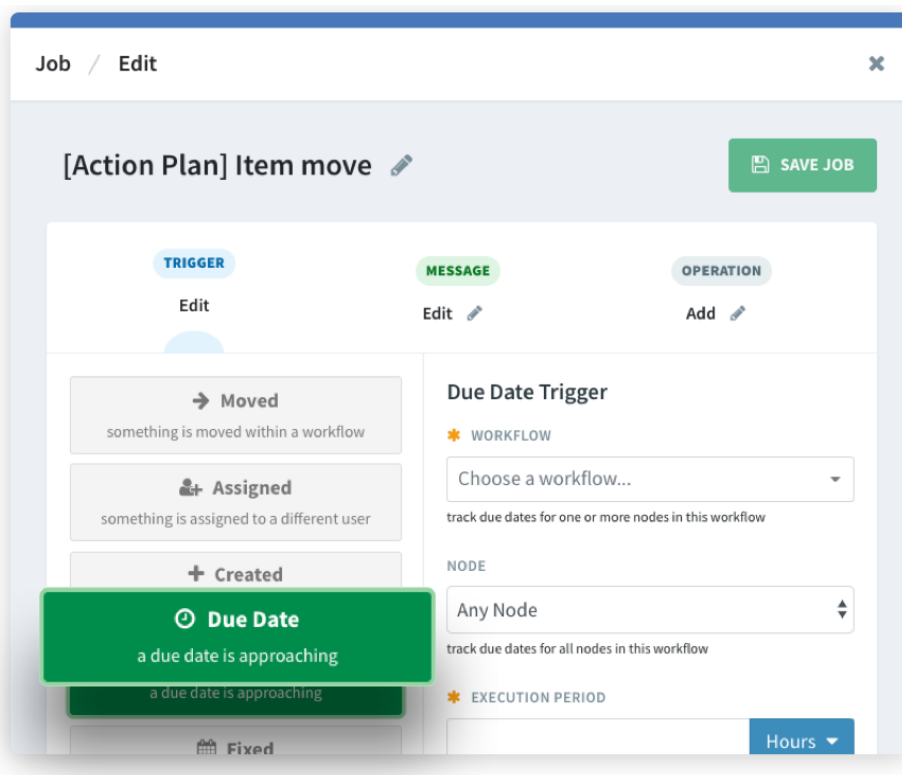


Рисунок 2 – Відстеження та пом'якшення ситуацій, пов'язаних із управлінням ризиками

Крім того, LogicGate Risk Cloud дає можливість точно визначити та оцінити ризики (Рисунок 3).



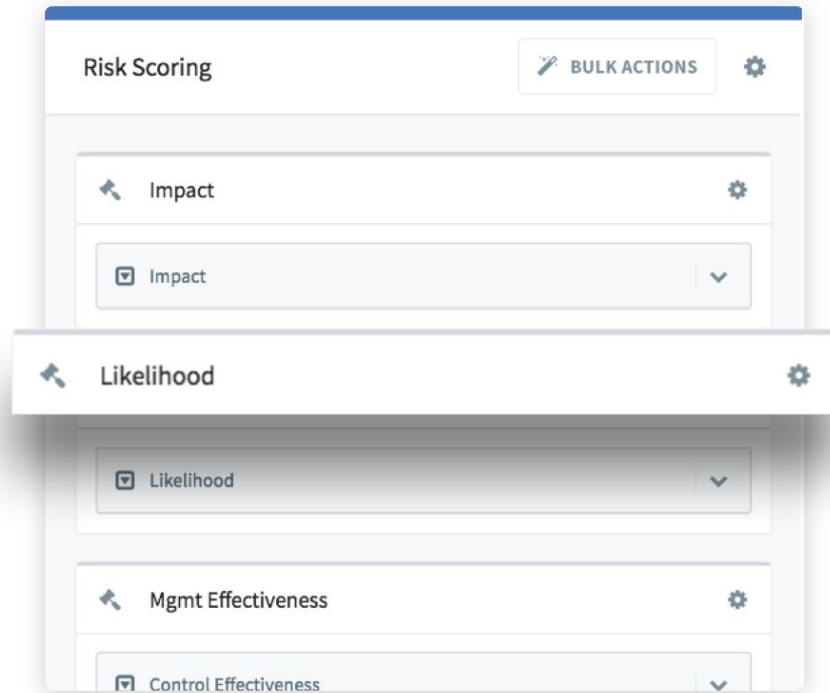


Рисунок 3 – Оцінка ризиків та управління ризиками підприємства

Продукт, запропонований LogicGate, є простим у використанні та досить ефективним для застосування у різних сферах бізнесу. Він надає можливість легко проектувати форми, полегшує управління робочим процесом та створення звітів. Великою перевагою є те, що користувач не обмежений набором процесів чи форм і може легко пристосувати будь-який крок (процес) до своїх конкретних потреб, залучити до роботи як внутрішніх, так і зовнішніх користувачів (через відкриті посилання). Таким чином, гнучкість, яку пропонує LogicGate, дозволяє розробляти процеси, що відображають структуру ризиків конкретної компанії. Надзвичайно різноманітна функціональність пропонується для збору та співставлення інформації про ризики. Система LogicGate корисна для реалізації рекурентних завдань та процесів. Інтерфейс інтуїтивно зрозумілий, проте у разі виникнення проблем можна скористатися функцією підтримки. Перевагою є наявність безлічі

варіантів реалізації різних процесів з можливістю залучення до роботи як внутрішніх, так і зовнішніх користувачів (через відкриті посилання).

Також на сайті представлено топ 8 безкоштовного програмного забезпечення для управління операційним ризиком у 2020 році:

1. Fusion Framework System.
2. DecisionTools Suite.
3. Hyperion.
4. Oplift.
5. Camms.Risk.
6. Dataminr.
7. Nvolve.
8. TenForce.

Розглянемо Fusion Framework System [18].

Система коротко Fusion Framework® узгоджує стратегічні цілі компанії з ключовими методами управління ризиками за допомогою гнучких інструментів Agile. Є можливість встановити відповідний контекст для аналізу, оцінки, моніторингу та реагування на ризики, виконати інтегрування власних даних на підприємстві для прийняття обґрунтованих рішень.

Fusion пропонує гнучкий набір інструментів, які адаптуються до програми клієнта. Незалежно від того, виконується керування цілою базою даних ризиків або програмою забезпечення безперервного ведення малого бізнесу, Систему Fusion можна налаштувати відповідно до конкретних потреб (Рисунок 4).

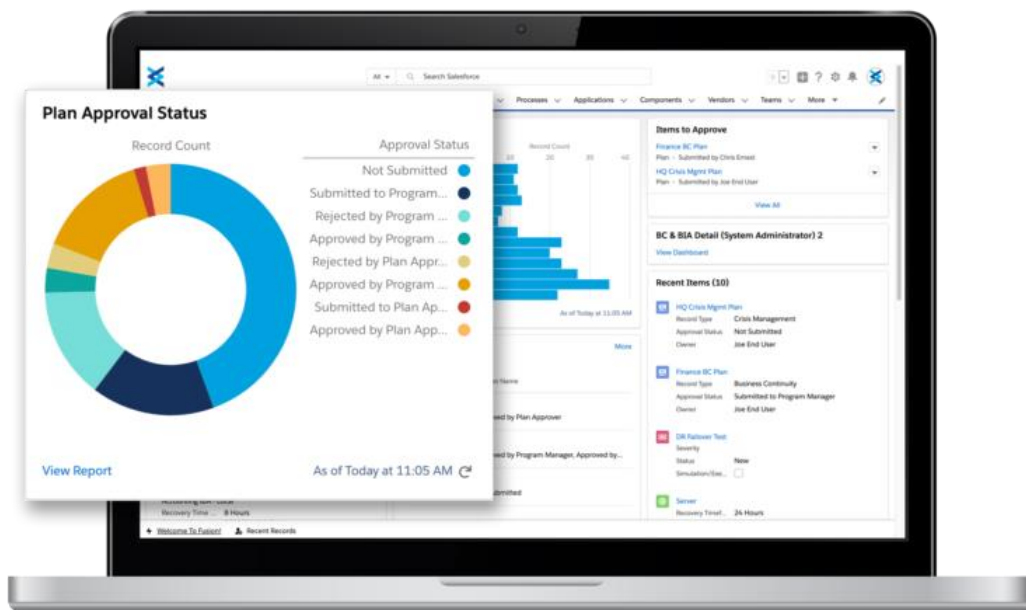


Рисунок 4 – Загальний вигляд інтерфейсу системи

Система Fusion дозволяє переглядати взаємозв'язки, які впливають на бізнес, на основі процесів і додатків (Рисунок 5).



Рисунок 5 – Візуальне відображення взаємозв'язків

Система створення звітів, а також наглядні і зручні у використанні інформаційні панелі дають можливість ефективно інформувати про найважливіші та найефективніші дані про ризики працівників всіх рівнів організації (Рисунок 6).

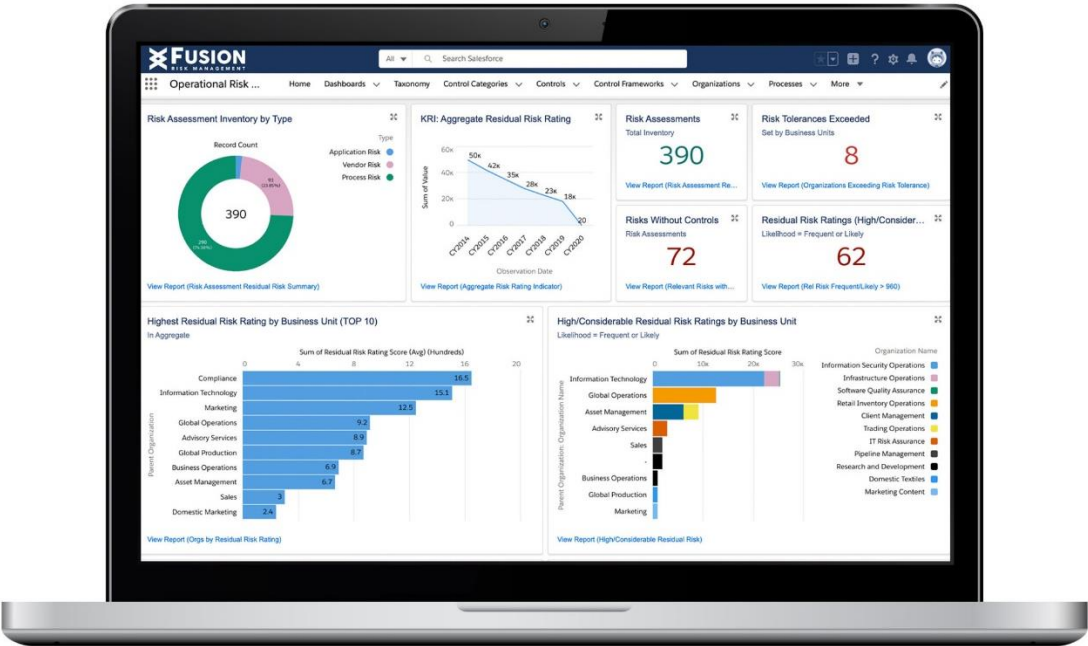


Рисунок 6 – Фонд спільної інформації

Крім цього, система дає можливість виявляти ризики і недоліки засобів контролю, визначати ризики та заходи контролю, а також оцінювати невід'ємний і залишковий ризики (Рисунок 7) .

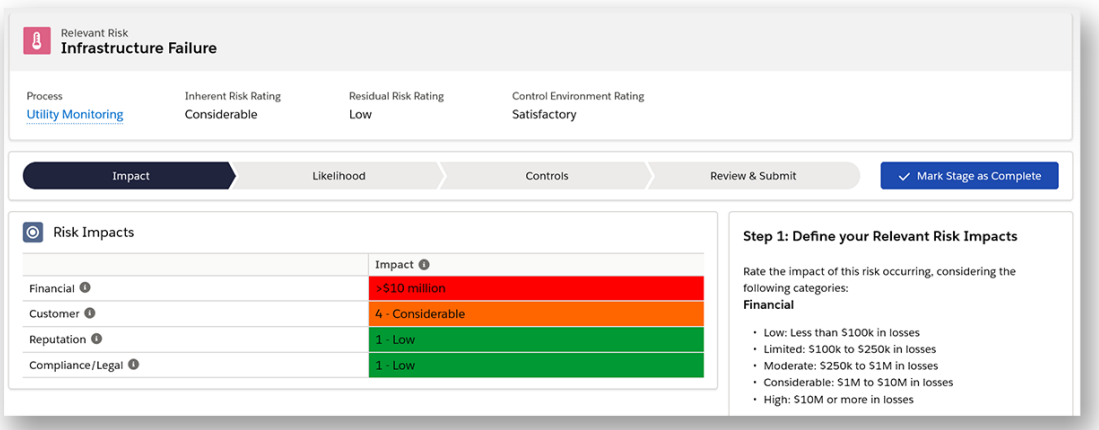


Рисунок 7 – Оцінка ризиків

Розглянемо програмне забезпечення SpheraCloud, описане на [19].

Sphera є провідним постачальником інтегрованого програмного забезпечення для управління ризиками, даних та консалтингових послуг, які фокусуються на навколишньому середовищі, охороні здоров'я, безпеці та стабільності (EHS & S), управлінні операційними ризиками та управлінні продуктами.

SpheraCloud – це мобільна платформа на базі SaaS, побудована на перевірених передових практиках, що поєднує ризики з усієї організації. Платформа SpheraCloud розроблена, щоб допомогти компаніям поліпшити процес прийняття рішень та продуктивність завдяки комплексному представленню їхнього портфеля ризиків (Рисунок 8).



Рисунок 8 – Рішення, доступні на SpheraCloud

Розглянемо детальніше рішення, які пропонуються для роботи із операційними ризиками.

Розширене програмне забезпечення та послуги з оцінювання ризиків дозволяють послідовно визначати, оцінювати та контролювати вплив операційного ризику. За допомогою розширеного програмного забезпечення для оцінки ризиків Sphera можна точно визначити ризик, оцінити його потенціал та зберегти контроль.

Заздалегідь побудовані та налаштовані шаблони надають користувачу інструменти, необхідні для виявлення прихованих ризиків та прогалин у безпеці.

Централізований доступ до бібліотек процесів, якості, безпеки та професійних знань та отриманих уроків дозволяє командам співпрацювати та визначати відстежуваність, забезпечувати відповідність та надійність контролю.

Уніфікований підхід до діяльності з управління ризиками та якістю забезпечує кращу відповідність чинним нормативним актам та законодавству про ризики.

Розширені засоби візуалізації забезпечують широкий доступ підприємств до спільного перегляду профілю ризику, послідовних показників для моніторингу ефективності контролю та загальної звітності про результативність.

Програмне забезпечення для контролю роботи дає змогу дізнатися, що відбувається, коли це відбувається, де це відбувається і що збільшує операційний ризик в режимі реального часу. Особливістю даного продукту є інтегрована візуалізація робочої діяльності та ризиків в режимі реального часу. Потужні можливості інтеграції дозволяють інтегруватися з існуючою екосистемою ІТ та бізнес-додатків для створення єдиного загального уявлення. Також користувач має можливість розглядати різні сценарії та робити висновки про операційний ризик.

Програмне забезпечення Sphera Master Data Management дає змогу глобально структурувати, візуалізувати і керувати своїми даними протягом усього їх життєвого циклу [19].

## 1.6 Висновки до розділу 1

Операційний ризик є надзвичайно важливою і актуальною проблемою для банків і фінансових установ в цілому [13].

Фірми зацікавлені в аналізі операційного ризику, оскільки його вплив може бути фатальним. А отже, операційний ризик є одним із основних ризиків фінансового сектору [15].

Нині проблеми у економіці мають навіть провідні держави світу, а отже, постає серйозна проблема зростання фінансових ризиків. На тлі теперішніх подій, коли безліч компаній і підприємств змушені призупиняти свою діяльність або взагалі виявляються банкрутами, важливою необхідністю є прийняття обґрунтованих управлінських рішень на основі високоякісних моделей ризиків і прогнозів розвитку фінансових процесів.

У зв'язку з поширенням COVID-19 посилився вплив проблем інформаційної безпеки, ведення документації та шахрайства, пов'язаного з кіберпростором, на роботу установ. Тому моделювання і оцінювання операційних ризиків є актуальною проблемою.

У даному розділі розглянуто поняття фінансового та операційного ризиків, їх класифікацію. Також подано огляд літератури та результатів відомих досліджень на дану тему. Крім цього наведено деякі існуючі методи моделювання та оцінювання операційного ризику. Було проведено детальний аналіз сучасного програмного забезпечення, яке вважається найпопулярнішим та найзручнішим для роботи із операційним ризиком, за яким Fusion Framework System є найкращою серед безкоштовних додатків, а LogicGate Risk Cloud – серед платних.

Отже, у розділі показано актуальність дослідження, на основі цього сформульовано постановку задачі.

#### Постановка задачі дослідження

1. Виконати огляд літератури, результатів відомих досліджень на тему, пов'язану з аналізом операційних ризиків.
2. Виконати огляд методів моделювання та оцінювання операційних ризиків.
3. Зібрати необхідні статистичні дані для виконання обчислювальних експериментів стосовно моделювання та оцінювання операційних ризиків.
4. Побудувати моделі операційних ризиків у формі мережі Байєса і дерев рішень, застосувати створені моделі для оцінювання операційного ризику.
5. Виконати аналіз отриманих результатів і зробити висновки; сформулювати перспективи подальших досліджень.



## **РОЗДІЛ 2 ВИБРАНІ МЕТОДИ І ПІДХОДИ ДО МОДЕЛЮВАННЯ І ОЦІНЮВАННЯ РИЗИКІВ ОПЕРАЦІЙНИХ ВТРАТ**

У цьому розділі подано поняття байєсівської мережі та їх класифікацію. Крім цього розглянуто методи побудови і оцінювання структури мереж Байєса, які включають методи на основі оціночних функцій, методи на основі використання тестів на умовну незалежність та інші. Також детально розглянуто процес формування імовірнісного висновку за алгоритмом Hugin, що включає моралізацію, тріангуляцію, визначення об'єднаного дерева та пропагацію (propagation) доказів. Крім цього розглянуто дерева рішень CART, випадковий ліс.

### **2.1 Ймовірнісний підхід – байєсівські мережі**

Серед ефективних методів моделювання і оцінювання операційних ризиків байєсівські мережі мають наступні переваги:

- вони дозволяють явно показати причинно-наслідкові зв'язки між факторами і відповідними операційними ризиками;
- дають можливість враховувати не лише безпосередні залежності рівня ризику від факторів ризику, а також залежності між факторами ризику;
- застосовуються до найрізноманітніших операційних ризиків.

Окрім категорій, де доступні дані дозволяють моделювати операційний ризик за допомогою стандартних статистичних моделей, БМ застосовуються у областях, де дані важче кількісно оцінити, наприклад, людські ризики;

- дозволяють працювати з неповними даними;
- дають змогу поєднати історичні дані та експертні оцінки;
- дають можливість виконувати аналіз сценаріїв для вимірювання максимальних операційних втрат.

### 2.1.1 Теорема Байєса та обчислення ймовірностей

Нехай події  $A$  та  $B$  – залежні. Їхня спільна ймовірність обчислюється наступним чином:

$$P(A, B) = P(A|B)P(B), \quad (2.1)$$

де  $A$  та  $B$  – залежні події.

Тобто ймовірність спільного виникнення подій  $A$  і  $B$  дорівнює ймовірності появи події  $A$ , за умови, що подія  $B$  вже відбулася, помноженій на ймовірність події  $B$ . Оскільки функція комутативна, можемо записати її у такому вигляді:

$$P(A, B) = P(B | A) P(A) \quad (2.2)$$

Прирівнявши праві частини цих виразів, маємо просту форму теорему Байєса:

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)} \quad (2.3)$$

Це можна інтерпретувати так: апостеріорна ймовірність  $P(B|A)$  дорівнює апіорній ймовірності  $P(B)$ , помноженій на відношення  $P(A|B) / P(A)$ .  $P(A|B)$  називається правдоподібністю.

Таким чином, комбінуючи суб'єктивну інформацію і емпіричні спостереження, отримуємо апостеріорну ймовірність.

Вважаючи  $\lambda = 1/P(A)$  нормуючою константою, можемо записати теорему Байєса у наступному вигляді:

$$P(B|A) = \lambda * P(A|B) P(B)$$

Важливим для побудови байєсівської мережі є припущення про вичерпність подій і про те, що події не перетинаються. Якщо ці умови виконуються,  $P(A)$  можна обчислити, використовуючи умовні ймовірності:

$$P(A) = \sum_{i=1}^n P(A, B_i) = \sum_{i=1}^n P(A|B_i) * P(B_i), \quad (2.4)$$

де  $B_i$  –  $i$ -те значення, якого набуває  $B$ ,

$n$  – кількість можливих значень  $B$ .

Підставивши (2.4) у формулу (2.3), отримаємо вираз:

$$P(B_i|A) = \frac{P(A|B_k) * P(B_k)}{\sum_{i=1}^n P(A|B_i) * P(B_i)}, \quad (2.5)$$

де  $B_k$  означає будь-яку гіпотезу з  $n$  можливих.

Ця формула є основою для побудови байєсівської мережі [27].

### 2.1.2 Означення мережі Байєса та класифікація

Мережа Байєса (МБ) – це трійка  $N = \langle V, G, J \rangle$ , де  $V$  – множина змінних;  $G$  – спрямований ациклічний граф, вузли якого відповідають випадковим змінним модельованого процесу;  $J$  – спільний розподіл ймовірностей змінних  $V = \{X_1, X_2, \dots, X_n\}$ . Виконується марковська умова: кожна змінна мережі не залежить від усіх інших змінних, за винятком батьківських попередників цієї змінної [20].

Як відомо, вузли байєсівської мережі з'єднані дугами, які вказують на існування причинного зв'язку між вузлами, тобто байєсівська мережа – це спрямований граф.

Детально про байєсівські мережі написано в [21].

Графічна структура мережі Байєса може бути трьох типів [21]:

- дерево – структура, у якій будь-яка вершина може мати не більше, ніж одну батьківську;
- полідерево – структура мережі Байєса, у якій будь-яка вершина може мати більше, ніж одну батьківську вершину, але при цьому між будь-якими двома вершинами повинно бути не більше одного з'єднуючого їх шляху;
- мережа – це структура, у якій будь-яка вершина може мати більш, ніж одну батьківську вершину; при цьому між будь-якими двома вершинами може бути більше одного з'єднуючого їх шляху.

Структури «дерево» та «полідерево» називають також однозв'язними мережами, а структури типу «мережа» – багатозв'язними мережами.

Спрямований ациклічний граф називають одиночно з'єднаним, якщо між будь-якими двома його змінними існує тільки один шлях, що їх з'єднує (дерева та

полі-дерева). Такі графи простіше аналізувати, проте вони рідко застосовуються до реальних процесів та подій.

Дочірній вузол мережі Байєса – вузол, що залежить від одного або більше інших вузлів. Батьківський вузол – вузол, що має хоча б одного вузла-нащадка. До множини нащадків відносять похідні вузли, які відносяться до одного батьківського вузла, а також дочірні вузли дочірніх вузлів вищого рівня [21].

Кореневий вузол – вузол, який не має жодного батьківського вузла.

Лист – вузол, що не має вузлів-нащадків.

Зв'язки між вершинами бувають трьох типів:

- лінійний (Рисунок 9а);
- розбіжний (divergent) (Рисунок 9б);
- збіжний (convergent) (Рисунок 9в).

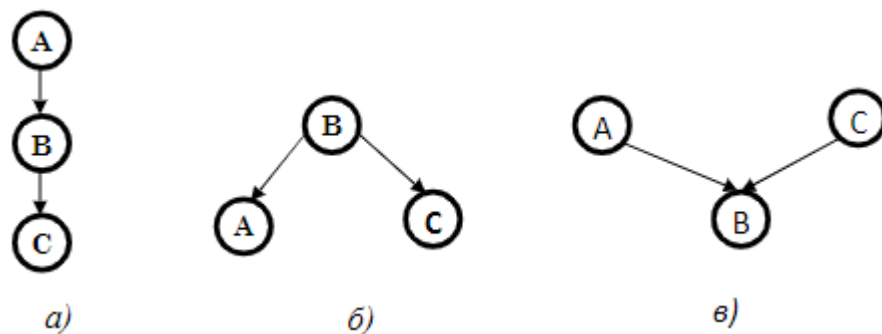


Рисунок 9 – Типи зв'язків між вершинами графу

Існують такі типи мереж Байєса [21]:

- дискретні – мережі, у яких змінні вузлів представлені дискретними величинами. У байєсівських мережах довіри вершини є випадковими змінними, дуги відображають ймовірнісні залежності, які визначаються через таблиці умовних ймовірностей. У таблиці умовних ймовірностей кожної вершини містяться ймовірності станів цієї вершини за умови станів її батьківських вузлів;

- динамічні – мережі, у яких значення вузлів змінюються з часом; використовуються для моделювання часових процесів;

- неперервні – мережі, в яких змінні вузлів – це неперервні величини; використовуються для моделювання стохастичних процесів у просторі станів з неперервним часом;
- гібридні – мережі, які містять вузли з дискретними і неперервними змінними.

### 2.1.3 Методи побудови і оцінювання структури мереж Байєса

Зазвичай, методи побудови і оцінювання структури мереж Байєса умовно поділяють на дві групи [22]:

- методи на основі оціночних функцій (search & scoring),
- методи на основі тесту на умовну незалежність (dependency analysis).

Основними проблемами багатьох методів є необхідність наявності упорядкованої множини вершин, що насправді при роботі з реальними даними може не виконуватися, та зменшення ефективності роботи методів, які використовують тест на умовну незалежність, у зв'язку із експоненційним зростанням об'єму обчислень. Існує також проблема побудови великих мереж Байєса із сотнями вершин [22].

У [22] зроблено досить детальний огляд цих методів. Далі розглянемо деякі з них.

### 2.1.3.1 Методи на основі оціночних функцій

Дерево Чу і Ліу. В основі алгоритму, запропонованого у 1968 році, лежить використання значень взаємної інформації між вершинами. У результаті застосування цього методу отримуємо структуру із значенням спільного розподілу ймовірностей мережі, який найбільше відповідає навчальним даним. Побудова структури БМ здійснюється за  $O(N^2)$  кроків, де  $N$  – кількість вершин мережі. Цей алгоритм не працює для багатозв'язаних БМ [22].

У 1988 році Рібаном і Перлом запропоновано удосконалений алгоритм Чу і Ліу для побудови БМ у вигляді полідерава [22].

К2. Розроблений Купером і Гершковичем в 1992 році відомий алгоритм, який виконує пошук структури з максимальним значенням функції Купера-Гершковича (КГ). Для роботи алгоритму потрібна наявність УМВ [22].

HGC. Запропонований у 1994 році. Суттєвою відмінністю HGC від інших алгоритмів, що ґрунтуються на оціночних функціях, є використання (вперше) параметричної модульності та рівнозначності подій. Одночасне застосування цих понять дозволяє об'єднувати статистичну інформацію та експертні знання для побудові БМ [22].

Алгоритм Лема-Бахуса. Запропонований у 1996 році, виконує евристичну побудову структури мережі, використовуючи значення взаємної інформації між вершинами, а як оціночна функція використовується функція опису мінімальною довжиною MDL (minimum description length) [22].

Алгоритм Бенедикта. Запропонований у 1996 році, виконує евристичний пошук на основі упорядкованої множини вершин, аналізуючи умовні

незалежності в структурі мережі на основі d-розділення, а як функція оцінки використовується ентропія [22].

Жадібний алгоритм пошуку та оцінки (Greedy Search-And-Score) [23]. В основі цього алгоритму лежить функція оцінки. Виконується пошук у просторі можливих мережевих структур, за допомогою функції оцінки визначається найкраща структура мережі-кандидата – та, що набрала найвищий бал. Під час пошуку за поточним кандидатом між вузлами дуги можуть додаватися, видалятися і змінювати напрям.

Мережева структура-кандидат забезпечується оцінкою максимальної правдоподібності таблиць умовної ймовірності, пов'язаних з вузлами мережі, щоб обчислити оцінку структури-кандидата. Припускається, що функція оцінки розкладається: загальний бал може бути обчислений як сума балів вузлів у мережі. Як функція оцінки може використовуватися інформаційний критерій Акайке (AIC) або інформаційний критерій Байєса (BIC).

Tree Augmented Naive Bayes (TAN) - це імовірнісна графічна модель, яка використовується для моделювання величезних наборів даних, що містять багато невизначеностей серед різних взаємозалежних наборів характеристик. У моделі TAN кожна змінна графу може мати двох батьків, а саме вузол класу та інший вузол змінної, за винятком однієї змінної, яка називається кореневою. Оскільки ця модель включає залежності між атрибутами, вона є більш реалістичною, ніж модель наївського Байєса. Модель TAN дає кращі результати роботи, якщо між атрибутами існує кореляція [24]. Процедура побудови цього дерева описана Чо та Ліу [25].



### 2.1.3.2 Методи на основі використання тестів на умовну незалежність

Алгоритм Вермута і Лоуренса. Запропонований у 1983 році, для побудови структури БМ застосовується тест на умовну незалежність. Алгоритмом виконується послідовний перебір упорядкованої множини вершин. Для кожної пари вершин  $X_i$  та  $X_j$ , таких, що  $X_i$  – предок  $X_j$ , обчислюється значення умовної незалежності. Цей алгоритм гарантує побудову БМ за навчальними даними, але недоліком є необхідність обчислення великої кількості тестів на умовну незалежність між вершинами. Тому цей алгоритм використовується лише для мереж із невеликою кількістю вершин [22].

Алгоритм побудови скінченного спрямованого ациклічного графа (boundary DAG algorithm). Запропонований Перлом у 1988 році. Для побудови БМ використовуються упорядкована множина вершин, функція спільного розподілу або достатньо велика навчальна вибірка даних. На відміну від алгоритму Вермута і Лоуренса, у цьому алгоритмі не потрібно проводити розрахунок великої кількості тестів на умовну незалежність, крім випадку застосування цього алгоритму для побудови марковських мереж [22].

SRA алгоритм. Запропонований у 1990 році, є модифікацією алгоритму скінченного спрямованого ациклічного графа. Для побудови БМ достатньо мати частково упорядковану множину вершин та ще деякі обмеження. Побудова БМ виконується послідовним додаванням дуг між вершинами з використанням евристичного пошуку. Недоліком алгоритму є необхідність обчислення експоненціальної кількості тестів на умовну незалежність [22].

SGS алгоритм. Запропонований у 1990 році, не потребує наявності упорядкованої множини вершин для побудови структури БМ, проте замість цього

необхідно виконувати експоненціальну кількість тестів на умовну незалежність між вершинами [22].

PC алгоритм. Розроблений у 1991 році, є удосконаленим варіантом SGS алгоритму. PC алгоритм застосовується для побудови мереж із невеликою кількістю дуг між вершинами (розріджених БМ) [22].

NPC (Necessary Path Condition). Розроблений дослідниками з Siemens. Основа для роботи у алгоритмів PC та NPC однакова. NPC є модифікацією PC, дозволяє працювати з обмеженими наборами даних [22].

### 2.1.3.3 Інші методи

Існують також й інші методи побудови структури БМ, які використовуються при неповноті даних (алгоритми стиснення границь), відсутності деяких змінних, які впливають на процес (приховані або латентні змінні) (алгоритми евристичного пошуку, які намагаються враховувати приховані змінні при моделюванні).

Для випадку, коли навчальні дані неповні або частина з них невірна також застосовуються алгоритми, які використовують значення максимального математичного очікування (ЕМ – expectation maximization).

ЕМ – алгоритм максимізації математичного очікування. Запропонований у 1977 році, намагається знайти локальні оптимальні оцінки максимальної правдоподібності параметрів. Головна ідея алгоритму полягає у тому, що за наявності значень усіх вузлів, навчання (на кроці  $M$ ) буде простим, оскільки наявна вся необхідна інформація. Тому на кроці  $E$  виконується обчислення значення очікуваної правдоподібності (expectation of likelihood), включаючи латентні змінні, так ніби вони спостерігались. На кроці  $M$  робиться обчислення значення максимальної правдоподібності параметрів, використовуючи

максимізацію значень очікуваної правдоподібності, отриманих на кроці E . Далі алгоритм знову виконує крок E з використанням параметрів, отриманих на кроці M, і так далі [22].

До інших методів можна також віднести побудову мережі Байєса на основі експертних суджень. Спочатку необхідно створити так звану карту ризиків і процесів, щоб визначити фактори ризику і причинність. Завдяки зіставленню ризиків та процесів отримуємо перелік видів дій, залежностей, можливих ризиків, причинних факторів та ключових індикаторів ризику, які можуть бути вузлами для мережі. З цього списку для використання в байєсівській мережі вибирається те, що є актуальним для конкретної проблеми, яка вирішується. Змінні, що описують одне і те ж, об'єднуються в один вузол із загальною назвою. Таким чином отримуємо вузли мережі та причинно-наслідкові зв'язки. Далі необхідно заповнити таблиці імовірностей.

Експертні судження можуть використовуватися для оцінки нових, рідкісних, складних або погано зрозумілих явищ або для прогнозування майбутніх подій. Процес збору та кодування експертних суджень за допомогою спеціальних методів усного або письмового спілкування називається виявленням ймовірності. Виявлення ймовірності вважається найскладнішим аспектом усього процесу побудови мережі Байєса [26].

Для спрощення цього процесу існують спеціальні методи, наприклад, внесення змін у графічну структуру мережі за допомогою розлучення батьків шляхом введення проміжних змінних або видалення дуг, що представляють слабкі залежності. Ці методи дають змогу зменшити кількість таблиць імовірностей, які потрібно заповнити експертам, та спростити їх.

Іншою складністю у роботі з експертними судженнями є включення психологічної складової.

Для вирішення цієї проблеми розроблені спеціальні протоколи. Трьома найпоширенішими протоколами, до яких схиляється більшість аналітиків, є

протокол Stanford/SRI, протоколи Моргана та Генріона та протокол Wallsten / EPA [26].

Протокол Stanford/SRI та протоколи Моргана та Генріон включають 5 етапів:

1. Мотивація експерта цілями процесу виявлення імовірностей.
2. Структурування невизначених величин однозначним способом.
3. Обґрунтування експертних суджень для уникнення когнітивних упереджень.
4. Кодування розподілів імовірностей.
5. Перевірка узгодженості отриманих розподілів.

Протокол Wallsten/EPA включає наступні етапи:

1. Підготовка документів, що окреслюють мету процесу.
2. Опис когнітивної евристики, упереджень та інших актуальних питань.

Експерти читають ці документи перед початком процесу виявлення імовірностей.

У процесі виявлення імовірностей можуть з'являтися упередження експертів. Упередженість – це аспект того, як люди вирішують проблеми. Воно може погіршити якість даних. У науковій літературі розрізняють когнітивне та мотиваційне упередження.

Мотиваційна упередженість зумовлена нашими людськими потребами, такими як соціальний тиск, бажане мислення серед інших тощо. Когнітивне упередження зумовлене способом обробки людиною інформації та виникає, коли оцінки експертів не відповідають нормативним статистичним або логічним правилам [26].

Тож після отримання імовірностей, визначених на основі експертних суджень, їх потрібно валідувати. Для цього перевіряється їх узгодженість та надійність.

Іншим методом є оцінка ефективності процесу виявлення імовірностей до проведення реального випробування, що дозволяє не проводити перевірку

отриманих імовірностей згодом. Цей підхід базується на припущенні, що ступінь близькості отриманих оцінок до того, що насправді вважає експерт, залежить від ефективності схеми виявлення імовірностей [26].

По завершенню виконання цих етапів маємо мережу Байєса із заповненими таблицями імовірностей для її вузлів.

#### 2.1.4 Формування імовірнісного висновку за алгоритмом Hugin

Формування ймовірнісного висновку - процес обчислення оцінки стану вершини на основі апріорної ймовірності про стани інших вершин мережі Байєса [20].

Алгоритм кластеризації Hugin запропонований у 1990 році Ф. Дженсеном, К. Олесеном, С. Андерсеном, є модифікацією LS-методу. Наведемо його опис, як зазначено у [27].

Формування висновку, зазвичай, виконується із включенням доказів (evidence), введених у вузли графа. Потім докази поширюються вздовж мережі до інших вузлів шляхом зворотного перегляду умовних розподілів за допомогою формування за теоремою Байєса спільних розподілів та встановлення маргінальних апостеріорних розподілів з отриманих розподілів. Для цього використовують об'єднані дерева – ненаправлені графи, що складаються з множини максимальних підграфів, відомих як кліки. Кліки – це групи вузлів, де кожен вузол у групі пов'язаний з кожним іншим вузлом у групі. Коли група вузлів з'єднана таким чином, вважається, що вони максимальні. Різні кліки в об'єднаному дереві з'єднані за допомогою сепараторів – спільних вузлів для двох сусідніх кліків.

Триангуляція – процес перетворення спрямованого ациклічного графа (DAG) у об'єднане дерево.

Процес виконується у три етапи:

- моралізація;
- триангуляція;
- визначення об'єданого дерева .

Моралізація мережі. На цьому кроці всі не пов'язані між собою батьківські вузли спільного дочірнього вузла спочатку об'єднуються ненаправленими ребрами. Решта направлених ребер робляться ненаправленими. У результаті отримаємо моралізований граф, оскільки всі батьки «одружені».

Триангуляція. Цикл – це послідовність вузлів, з'єднаних ребрами, які починаються і закінчуються в одному вузлі. Цикл довжиною  $n$  складається із  $n$  послідовних ребер.

Триангуляція – це процес додавання неорієнтованих ребер, таким чином, щоб будь-який цикл, що має довжину більше 3, містив акорд, де акорд визначається як ненаправлене ребро, що з'єднує два непослідовних вузли у розглянутому циклі.

Після того, як було отримано триангульований граф, можна визначити об'єднане дерево. Це передбачає ідентифікацію у графі клік та сепараторів, які їх з'єднують. Коли граф представлений у такій формі, докази можуть бути включені локально в кліки, де розрахунки мають менші розміри. Потім інформація з оновлених клік поступово поширюється до інших клік графіка через сепаратори. Таким чином, немає необхідності мати справу з великими спільними розподілами цілого графа.

Розглянемо процес отримання висновку по об'єданому дереві.

Основними кроками, які слід зробити для висновку за допомогою об'єднаного дерева, є:

- визначення апіорних розподілів;
- ініціалізація дерева;
- двофазне пропагація.

Для побудови БМ необхідно визначити апіорні розподіли змінних: безумовний розподіл для вузлів без батьків та умовний розподіл для дочірніх вузлів.

Вони можуть бути визначені на основі:

- суб'єктивного висновку експерта. Для отримання експертних оцінок проводиться анкетування, що дає змогу кількісно оцінити ймовірності. Очевидно, що для застосування цього методу необхідна достатня впевненість у точності порад експерта;
- оцінки максимальної правдоподібності. За цим методом, умовні апіорні ймовірності знаходяться за відношенням частоти події до частоти батьківської конфігурації. Наприклад, нам знадобляться безумовні апіорні  $P(A)$ ,  $P(B)$ ,  $P(E)$ ,  $P(F)$  і умовна апіорна  $P(C|A,B)$ . Нехай змінні можуть приймати по два стани – з індексами у і n. Для кожної апіорної ймовірності нам будуть потрібні ймовірності для кожної конфігурації комбінації станів задіяних змінних. Таким чином, для визначення  $P(C|A,B)$  нам будуть потрібні значення  $P(C_y|A_y, B_y)$ ,  $P(C_n|A_y, B_y)$ ,  $P(C_y|A_n, B_y)$  і так далі.

Отже, максимальну правдоподібність можна оцінити так:

$$P(C_y|A_y, B_y) = \frac{n(C_y, A_y, B_y)}{n(A_y, B_y)},$$

де  $C_y, A_y, B_y$  - відповідні стани змінних  $C, A, B$ .

Для безумовних апіорних імовірностей це буде просто частка пропорційності між різними станами змінної.

Спільна щільність по набору змінних  $U$ , що складається із змінних, позначених вектором  $X$ , може бути виражена як добуток умовних щільностей на кожному вузлі:

$$P(U) = \prod_X P(X|pa(X)). \quad (2.6)$$

Далі потрібно ініціалізувати дерево. Для кожної кліки та сепаратора існує таблиця з клітинками, які відповідають кожній комбінації станів змінних у кліці або сепараторі. Наприклад, кліка  $(A, B, C)$  матиме 8 комірок для комбінацій  $(A_u, B_u, C_u)$ ,  $(A_n, B_u, C_u)$ ,  $(A_n, B_n, C_u)$  тощо. Числові значення в цих клітинках називаються потенціалами. Вони змінюватимуться в міру того, як інформація передається уздовж об'єднаного дерева та в міру оновлення кожної кліки або сепаратора. Для ініціалізації всі потенціали встановлюються на одиницю. Потім апіорні імовірності, які розкладаються на множники відповідно до (2.6), множаться на кліки, що містять змінні у факторизованій множині пріоритетів.

Деякі кліки можуть мати більше змінних, ніж будь-який доступний набір факторизованих апіорних імовірностей. У цих випадках апіорні все ще множаться на кліки, якщо змінні в апіорах є підмножиною змінних кліків. Щодо кліків, для яких ця операція неможлива, усі потенціали залишаються одиницями. Потенціали в сепараторах також залишаються одиницями. Як тільки це буде зроблено, дерево можна використовувати для пропагації.

Пропагація (розповсюдження) включає повідомлення, що походять від листків дерева (тобто кліки на кінцях дерева) і передаються вздовж дерева до кореневої кліки, а потім знову назад до листя (вибір кореневої кліки довільний). По дорозі потенціали кожної кліки будуть оновлюватися. Отже, цей процес іноді називають двофазною пропагацією. Після завершення обох фаз дерево досягло



рівноваги. На цьому етапі потенціали в кожній кліці потрібно нормалізувати. Потім вони можуть бути маргіналізовані, щоб отримати апостеріорні маргінальні розподіли для кожної змінної вузла.

## 2.2 Древа рішень як альтернативний метод моделювання та оцінювання ризиків операційних втрат

Короткий огляд літератури [28] дає змогу стверджувати, що методи машинного навчання лише набирають популярності у сфері операційних ризиків. З розвитком технологій все більші обсяги даних необхідно збирати і обробляти щодня. Існує також проблема неструктурованості даних. Тож наразі проводяться активні дослідження застосування дерев рішення у фінансовій сфері, зокрема і для управління ризиками, адже вони дозволяють аналізувати та обробляти великі обсяги даних.

Застосування машинного навчання для роботи із операційними ризиками дозволяє зменшити ризик, тобто виявити та/або запобігти йому. У сфері операційного ризику, крім випадків кібербезпеки, машинне навчання переважно зосереджується на проблемах, пов'язаних з виявленням шахрайства та підозрілих транзакцій.

### 2.2.1 Поняття дерева рішень та їх класифікація

Класифікаційні та регресійні дерева (Classification and Regression Trees, CART) – це термін, введений Лео Брейманом для позначення алгоритмів дерев

рішень, які можуть бути використані для класифікаційних або регресійних проблем моделювання прогнозу [37].

Більш сучасною назвою алгоритму «дерева рішень» є термін CART, що використовується на деяких платформах.

Дерева рішень, що використовуються у інтелектуальному аналізі даних, поділяються на два основних типи:

- дерева класифікації – дерева, що вирішують проблему передбачення категоріальної цільової змінної;
- дерева регресії – дерева, що вирішують проблему передбачення неперервної цільової змінної.

Дерева, які використовуються для регресії, і дерева, які використовуються для класифікації, де у чому схожі, проте мають і відмінності, такі як процедура, яка використовується для визначення місця розділення [29].

Дерево рішень репрезентує процес класифікації. Припустимо, що всі змінні мають скінченну кількість дискретних значень, яких можуть набувати, та існує одна цільова змінна, для якої і потрібно виконати класифікацію. Значення, які цільова змінна може набувати, називаються класами. Дерево рішень - це дерево, в якому кожен нелистовий вузол позначений вхідною змінною. Дуги, спрямовані від цих вузлів, позначаються кожним із можливих значень цільової змінної або ведуть до підлеглих вузлів, позначених іншими вхідними змінними. Кожен лист дерева позначений класом або розподілом ймовірностей за класами цільової змінної [29].

Дерево будується шляхом розбиття даних, що містяться у кореневому вузлі дерева, на підмножини, що становлять дочірні вузли. Розбиття базується на наборі правил розділення. Процес ділення виконується рекурсивно, доки підмножина у вузлі не міститиме однакові значення цільової змінної (однорідні вузли). Процес поділу на дочірні вузли визначає якість роботи моделі [29].

### 2.2.2 Способи розділення вузлів

Способи виконання поділу на дочірні вузли можна розділити на дві категорії залежно від типу цільової змінної [30]:

а) для неперервної цільової змінної:

- зменшення дисперсії;

б) для категоріальної цільової змінної:

- домішка Джині (Gini Impurity);
- збільшення інформації (Information Gain);
- Chi-Square.

Зменшення варіації – це метод розділення вузла, у якому дисперсія використовується як міра для вирішення ознаки, за якою вузол поділяється на дочірні вузли. Дисперсія використовується для обчислення однорідності вузла і дорівнює нулю для повністю однорідного.

Для поділу дерева рішень за допомогою зменшення дисперсії необхідно виконати наступні кроки [30]:

1. Для кожного можливого поділу окремо обчисліть дисперсію кожного дочірнього вузла.
2. Обчисліть дисперсію кожного розбиття як середньозважену дисперсію дочірніх вузлів.
3. Виберіть варіант поділу, що має найменшу дисперсію.
4. Виконуйте кроки 1-3 до досягнення повністю однорідних вузлів.

Gini Impurity – це метод розділення вузлів для дерев рішень із категоріальною цільовою змінною. Це найпопулярніший і простий спосіб розбиття дерев рішень. За допомогою Gini Impurity можна оцінити «чистоту»

вузла, тобто якщо у вузлі містяться лише змінні із однаковим значенням класу, то Gini Impurity для цього вузла дорівнюватиме нулю.

Для поділу дерева рішень за допомогою домішки Джині необхідно виконати наступні кроки [30]:

1. Для кожного можливого поділу окремо обчисліть домішку Джині кожного дочірнього вузла.
2. Обчисліть домішку Джині кожного розбиття як середньозважену домішку Джині дочірніх вузлів.
3. Виберіть варіант поділу, що має найменше значення домішки Джині.
4. Повторюйте кроки 1-3, доки не буде отримано однорідних вузлів.

Метод збільшення інформації працює на основі використання поняття ентропії і визначається як одиниця мінус значення ентропії. Ентропія використовується для розрахунку чистоти вузла: чим менше значення ентропії, тим більша чистота вузла, ентропія однорідного вузла дорівнює нулю. Ентропія обчислюється за формулою:

$$Entropy = - \sum_{i=1}^n p_i \log_2 p_i,$$

де  $p_i$  – частотна імовірність  $i$ -го класу у даних (кількість записів із цільовою змінною  $i$ -го класу поділити на загальну кількість записів у датасеті) [30].

Алгоритм використання збільшення інформації містить такі ж кроки, як і алгоритм використання значення домішки Джині з використанням збільшення інформації відповідно.

### 2.2.3 Переваги і недоліки дерев рішень

Застосування дерев рішень має наступні переваги та недоліки [29].

Перевагами застосування CART є:

- дерева рішень можна застосовувати для передбачення змінної із багатьма класами;
- результати застосування CART легко інтерпретувати та розуміти, оскільки вони являють собою ряд умов «якщо-то»;
- метод застосовний як до числових, так і категоріальних даних;
- нелінійні взаємозв'язки між змінними не впливають на продуктивність дерев рішень;
- CART не вимагає жодної спеціальної підготовки даних, окрім як належного репрезентування проблеми.

Недоліками CART є:

- невелика зміна у наборі даних може призвести до значних змін у структурі дерева рішень, що робить його нестабільним;
- менш ефективні у передбаченні неперервних змінних, адже дерева рішень, як правило, втрачають інформацію, класифікуючи змінні за кількома категоріями.

### 2.2.4 Випадковий ліс

Деякі прийоми, які часто називають ансамблевими методами, будують більше одного дерева рішень. Одним із найпопулярніших таких алгоритмів є випадковий ліс (Random Forest). Даний алгоритм відноситься до так званих «bagging» алгоритмів та полягає у побудові у ході навчання численних дерев рішень, що продукують моду для класів у задачі класифікації або усереднений прогноз для регресії (Рисунок 10) [31]:

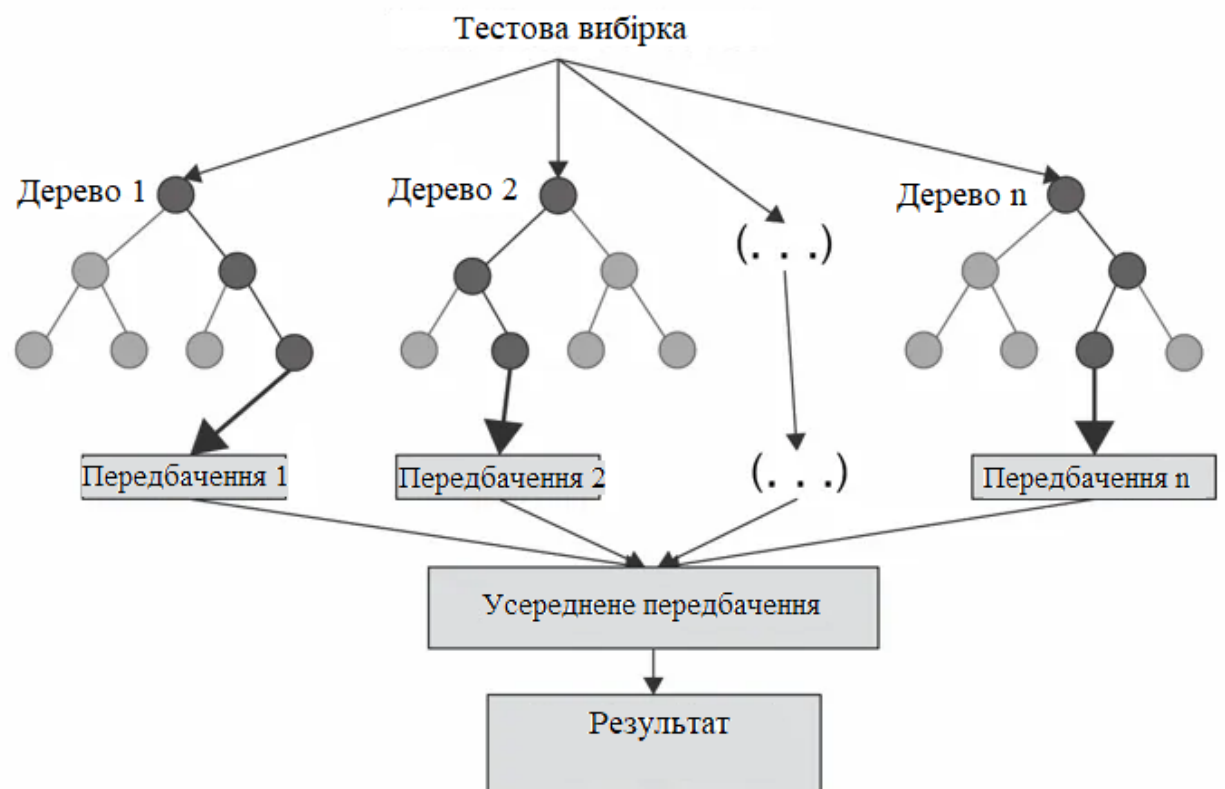


Рисунок 10 – Схема функціонування випадкового лісу

Наведемо алгоритм навчання класифікатора [32].

Припустимо, у навчальній вибірці міститься  $N$  записів, кількість ознак (змінних) становить  $M$ , і заданий параметр  $m \approx \sqrt{M}$ .

Усі дерева комітету будуються незалежно один від одного за такою процедурою:

1. Згенеруємо випадкову підвибірку з повторенням розміром  $n$  з навчальної вибірки.
2. Побудуємо дерево рішень, яке класифікує приклади даної підвибірки, причому в ході створення чергового вузла дерева будемо вибирати ознаку, на основі якої проводиться розбиття, не з усіх  $M$  ознак, а лише з  $m$  випадково вибраних. Вибір найкращої з  $m$  ознак здійснюватимемо за критерієм Джині.
3. Дерево будується до повного вичерпання підвибірки і не піддається процедурі відсікання, на відміну від дерев рішень CART.

Класифікація об'єктів проводиться шляхом голосування: кожне дерево комітету відносить об'єкт, який класифікується, до одного з класів, і перемагає клас, за який проголосувало найбільше число дерев.

Оптимальне число дерев підбирається таким чином, щоб мінімізувати помилку класифікатора на тестовій вибірці.

Застосування випадкового лісу має наступні переваги та недоліки [32].

Перевагами використання випадкового лісу є:

- здатність ефективно обробляти дані з великою кількістю змінних і класів цільової змінної;
- нечутливість до масштабування значень змінних;
- може використовуватися для категоріальних і неперервних даних;
- здатність працювати паралельно в багато потоків;
- масштабованість.

Недоліками випадкового лісу є:

- Схильність алгоритму до перенавчання;
- великий розмір отримуваних моделей.

### 2.3 Висновки до розділу 2

У другому розділі подано теорему Байєса, означення мереж Байєса, їх класифікацію. Крім цього, зроблено огляд методів побудови і оцінювання (навчання) структури мереж Байєса, детально розглянуто алгоритм Hugin для формування імовірнісного висновку за допомогою мережі Байєса. Також розглянуто дерева рішень CART та випадковий ліс, наведено їх переваги та недоліки стосовно практичного застосування.



## РОЗДІЛ 3 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ

У даному розділі виконується моделювання та оцінювання операційного ризику за допомогою мережі Байєса та альтернативного методу – дерев рішень. Результати роботи проілюстровані необхідними таблицями і рисунками.

### 3.1 Обґрунтування вибору платформи

Для побудови байєсівської мережі було обрано програмне середовище Nugin Lite. Воно надає можливість будувати байєсівські мережі, а також аналізувати їх. Зокрема наявні різні алгоритми навчання, дослідження параметрів та структури мережі, є можливість використання неперервних змінних, введення доказів у мережу та їх розповсюдження. Також перевагою даного середовища є зручний та зрозумілий інтерфейс, що полегшує роботу, а також його безкоштовність. Обмеження на кількість вузлів у мережі – 100 штук.

Таким чином, дане середовище не потребує глибоких знань та тривалого досвіду роботи, є доступним, зрозумілим і не займає значного об'єму пам'яті пристрою.

Для реалізації дерев рішень обрано програмне середовище Jupiter Notebook та мова програмування Python. Дана мова програмування має велику кількість вбудованих бібліотек, що суттєво пришвидшує роботу над програмою та дозволяє зосередитися саме на дослідженні застосування даного методу та зекономити час

на написанні загальнопоширених функцій і методів. Середовище Jupiter Notebook є безкоштовним і зручним для роботи з мовою програмування Python.

### 3.2 Вимоги до обладнання та інструменти для роботи з даними

Моделювання та оцінювання операційних ризиків було виконано за допомогою Hugin Lite.

Для роботи із створеною моделлю необхідне таке середовище:

- операційна система (Windows, MacOS, Linux);
- встановлений Hugin Lite 8.8;
- мережа у вигляді файлу з розширенням .oobn.

Для реалізації дерева рішень використовувалися вбудовані бібліотеки мови програмування Python, які наведено у таблиці 1.

Таблиця 1 – Використані бібліотеки Python

Назва бібліотеки	Версія	Призначення бібліотеки
pandas	0.25.3	Аналіз даних, статистика
numpy	1.18.4	Робота з масивами
imblearn	0.5.0	Збалансування даних
sklearn	0.0	Множина модулів для машинного навчання та аналізу даних
matplotlib	2.2.2	Графічне представлення даних

### 3.3 Опис архітектури СППР і функціональної схеми

Функціональна блок-схема архітектури СППР оцінювання капіталу, необхідного на покриття операційних ризиків, яку будуватимемо у даному розділі, зображена на Рисунку 11.

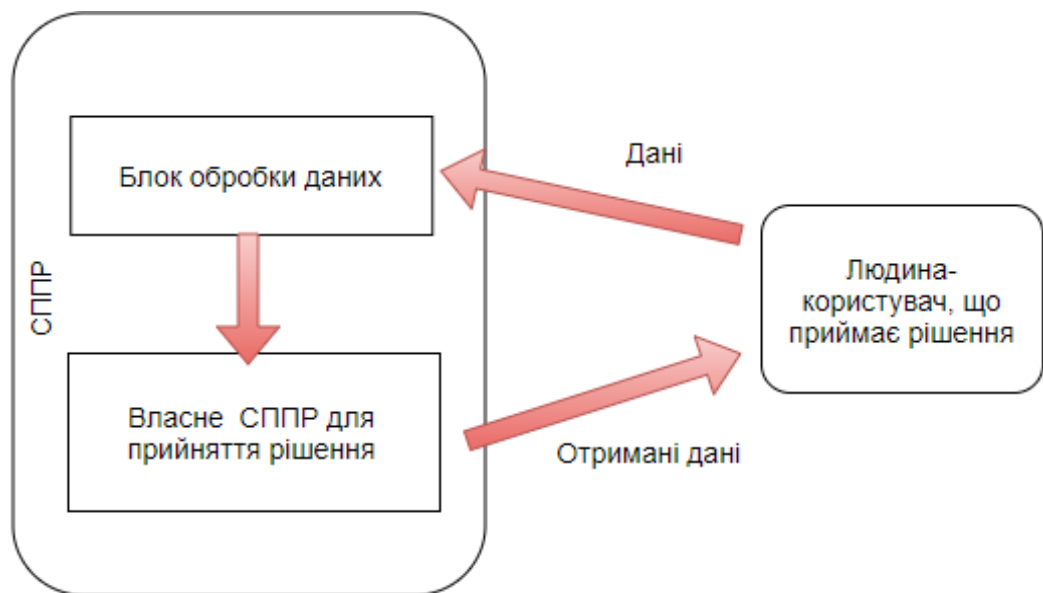


Рисунок 11 – Функціональна блок-схема архітектури СППР

Загальна робота даної системи може бути описана наступним чином:

- користувач вводить докази, які репрезентують поточний стан установи, у вузли мережі;
- виконується пропagaція цих доказів;
- на основі цих даних система розраховує імовірності потрапляння втрат у кожен проміжок і виводить їх користувачеві;

- користувач може продовжити роботу з системою, виконуючи сценарний аналіз;
- на основі отриманих ймовірностей користувач приймає рішення про подальші дії.

### 3.4 Побудова мережі Байєса для оцінювання операційних ризиків

Моделювання причинно-наслідкових зв'язків та оцінювання необхідного капіталу на покриття операційних ризиків продемонстровано на прикладі гіпотетичного комерційного банку.

#### 3.4.1 Визначення структури мережі

Побудову структури мережі Байєса для оцінювання операційного ризику було виконано на основі експертних суджень. З їх допомогою було обрано вузли – змінні мережі та встановлено причинно-наслідкові зв'язки між ними.

Так, для відображення процесів, притаманних даному комерційному банку, було обрано 16 вершин, назви яких, а також значення, які можуть набувати ці змінні, подано у таблиці 2.

Таблиця 2 – Вузли мережі, їх позначення і можливі значення

№	Назва вершини	Позначення у мережі	Значення, які може набувати змінна
1	Втрати	Loss	0K, 1K, 10K, 50K, 100K
2	Ризик зовнішнього середовища	Risk of external environment	Low, Medium, High
3	Хакерські атаки	Hacker attacks	Yes, No
4	Надійна система шифрування даних, використання Firewall	Data encryption Firewall	Yes, No
5	Зміни в законодавстві	Changes in legislation	Affect, no affect
6	Надійність платіжних систем та банків партнерів	Reliability of payment systems and partner banks	Low, Medium, High
7	Ризик персоналу	Risk of staff	Low, Medium, High
8	Кваліфікованість персоналу	Qualification of staff	Low, Medium, High
9	Можливість внутрішнього шахрайства	Internal fraud	Low, Medium, High
10	Моніторинг персоналу	Staff monitoring	Yes, No
11	Ризик системи (технологічний)	Technological risk	Low, Medium, High
12	Збої серверів	Server failure	Yes, No
13	Збої мережі	Network failures	Yes, No
14	Ризик процесу	Risk of process	Low, Medium, High
15	Використання автоматизації процесів	Using process automation	Yes, No/Low
16	Актуальна база бізнес-процесів	Current database of business processes	Yes, No

Як зазначено у [5], більшість українських банків дотримуються такої системи класифікації операційних ризиків:

- ризик персоналу – ризик, пов'язаний з діями працівників і безпекою робочого місця, тобто із можливими помилками співробітників, шахрайством, недостатньою кваліфікацією персоналу, можливістю несприятливих змін у трудовому законодавстві, травмуванням на робочому місці тощо;
- ризик систем і технологій – ризик, пов'язаний з перебоями і відмовами інформаційних систем, програм або баз даних, систем передачі інформації та обладнання, необхідного для діяльності банку;
- ризик процесу – ризик, пов'язаний із неузгодженими або неправильними бізнес процесами, методологічними помилками або неузгодженістю документів, відсутністю повної нормативної бази, недостатністю внутрішнього контролю, неузгодженістю дій при здійсненні операцій тощо;
- ризик зовнішнього середовища – ризик, пов'язаний із діями третіх осіб, зокрема зовнішнє шахрайство, а також ризик втрат в наслідок природних катастроф та інших фізичних впливів, які можуть нанести збитки основному устаткуванню, системам, технологіям і ресурсам банку.

Таким чином, цільовим вузлом є Loss (втрати). На нього впливають чотири види операційного ризику, які в свою чергу пов'язані із іншими причинами виникнення операційного ризику, визначеними для конкретного комерційного банку. Так, наприклад, хакерські атаки можуть стати причиною збою мережі, а також виникнення ризику зовнішнього середовища. Проте для протидії втручання зловмисників використовується фаєрвол та надійна система шифрування даних. Внутрішнє шахрайство спричиняє ризик персоналу, проте у банківській практиці рекомендують проводити моніторинг працівників для виявлення та запобігання можливості незаконних дій співробітників.

Отже, після визначення вузлів мережі та причинно-наслідкових зв'язків між ними було побудовано мережу Байєса (Рисунок 12).

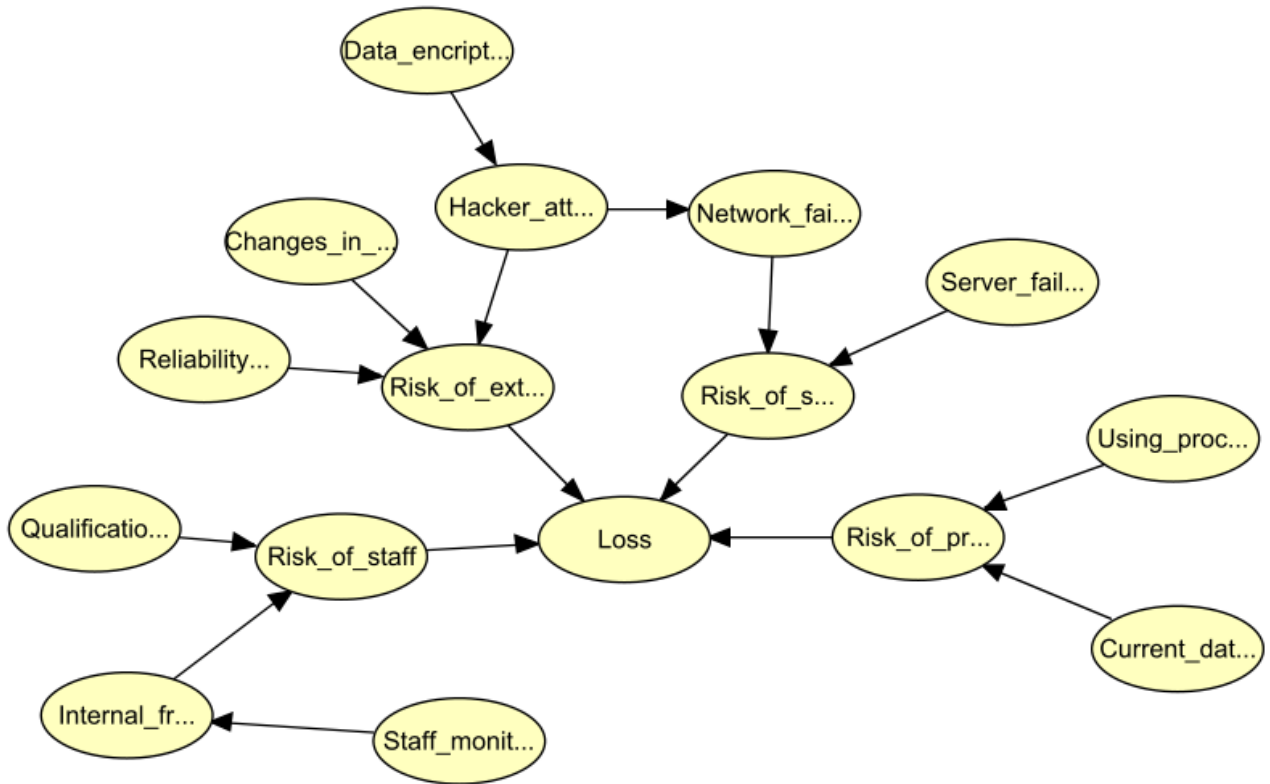


Рисунок 12 – Мережа Байєса для оцінювання операційного ризику комерційного банку

Валідація структури мережі була проведена експертами. Існує вбудоване припущення про незалежність, оскільки мережа була змодельована з точки зору причинності. Оскільки батьківські вузли мережі були обрані як безпосередні причини їх дочірніх вузлів, локальні умови умовної незалежності виконуються автоматично [33].

### 3.4.2 Визначення апіорних імовірностей

Після побудови структури мережі необхідно визначити апіорні імовірності вузлів-змінних. У даному випадку для визначення безумовних апіорних імовірностей вузлів-листіків мережі та умовних імовірностей інших вузлів було використано експертні судження. Таким чином було отримано таблиці імовірностей, подані у додатку Б. Апіорні маргінальні розподіли показано на Рисунку 13.

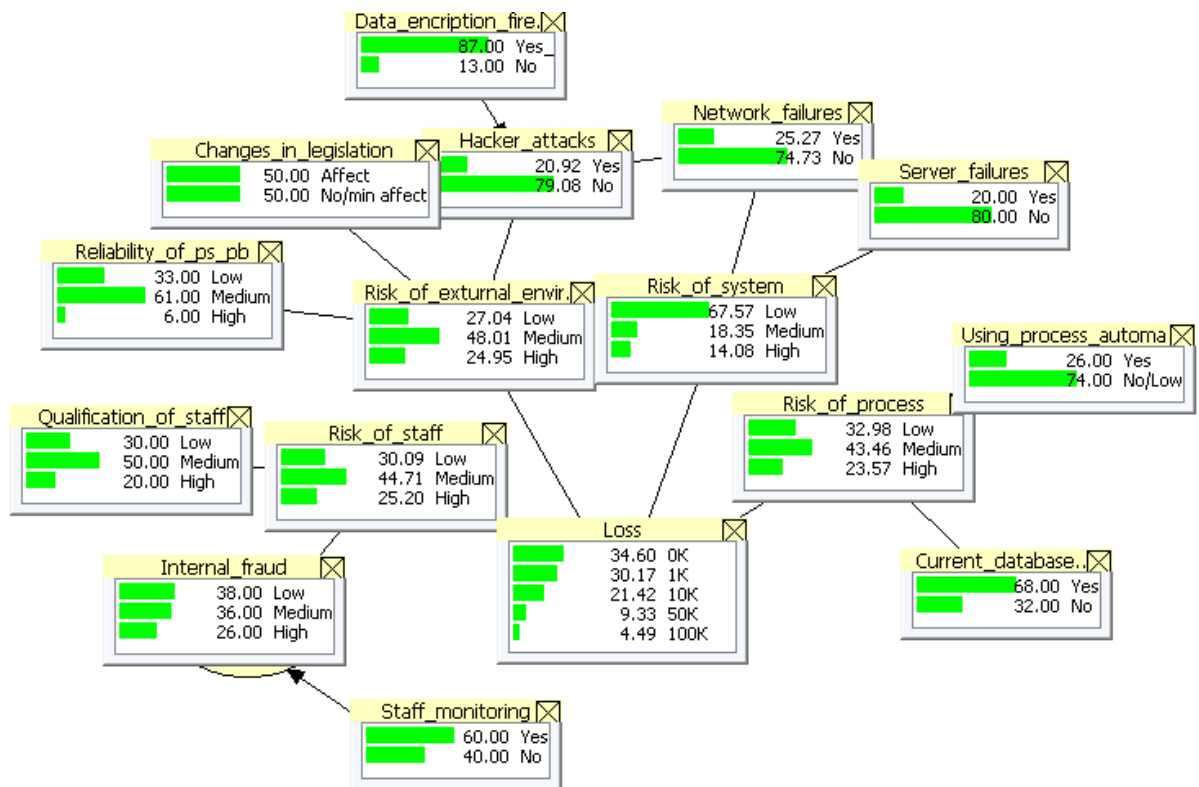


Рисунок 13 – Апіорні маргінальні розподіли



### 3.4.3 Обчислення необхідного капіталу на покриття ОР та сценарний аналіз

За допомогою побудованої мережі можемо оцінити необхідний капітал для покриття операційного ризику даного комерційного банку. Втрати рахуємо у тисячах доларів, для даного банку маємо наступну дискретизацію: 0К, 1К, 10К, 50К, 100К.

Отже, маємо наступний розподіл втрат під час виникнення операційного ризику у комерційному банку (таблиця 3).

Таблиця 3 – Розподіл втрат при виникненні ОР

Втрати	0К	1К	10К	50К	100К
Імовірність	0,3460	0,3017	0,2142	0,0933	0,0449

За даним розподілом втрат, використавши лінійну інтерполяцію, можемо обчислити 95-й центиль. Таким чином, необхідний капітал для покриття операційного ризику становить 47771 долар.

Припустимо, поточну ситуацію у даному банку відображає множина доказів (свідчень), наведена у таблиці 4.

Таблиця 4 – Введені докази

Змінна	Доказ
Data encryption Firewall	Yes
Reliability of payment systems and partner banks	Medium
Qualification of staff	Medium
Staff monitoring	No
Current database of business processes	No

Вводимо дані докази у мережу, виконуємо пропагацію і отримуємо апостеріорний розподіл втрат у наслідок виникнення операційного ризику (Рисунок 14).

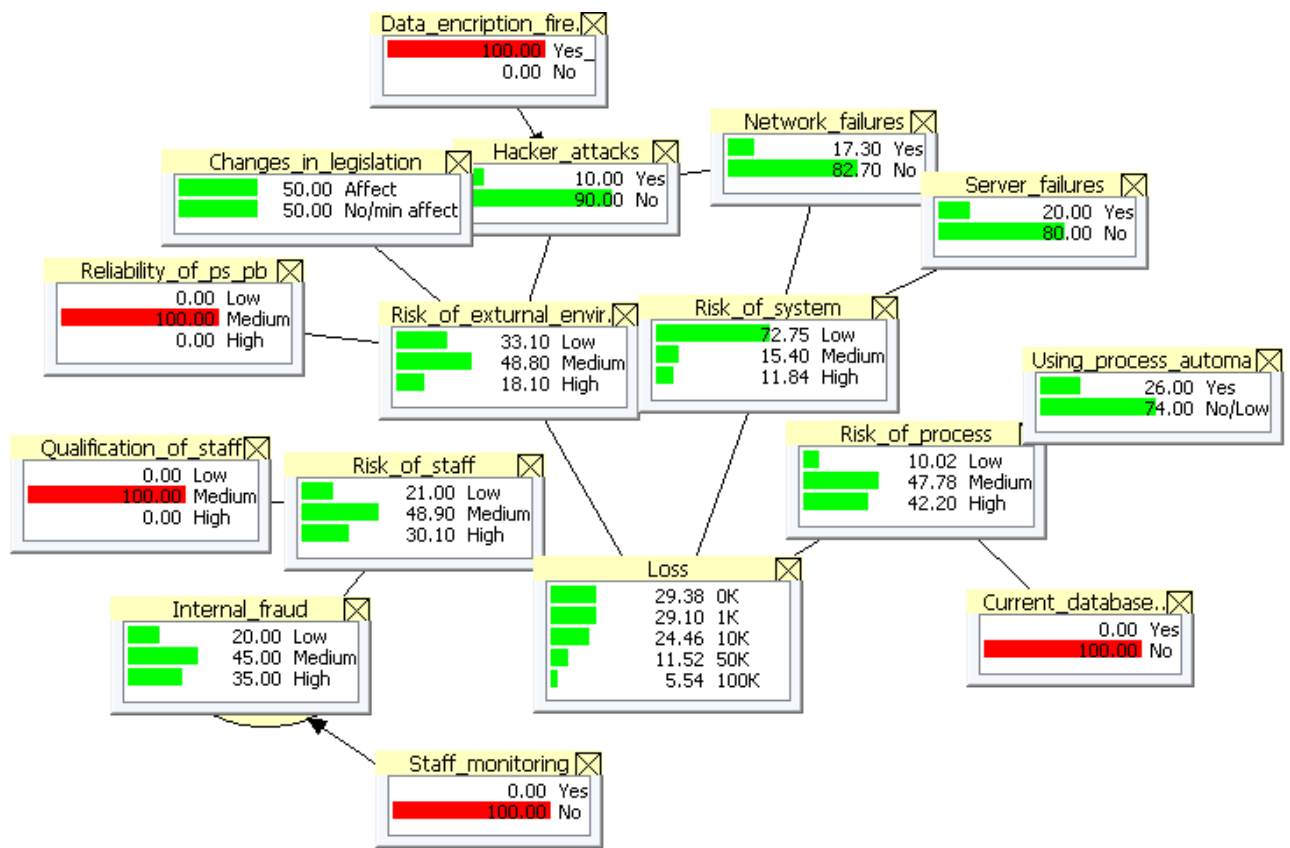


Рисунок 14 – Апостеріорні імовірності після введення доказів

Капітал для покриття втрат у цьому випадку становитиме 54874 долари.

Припустимо, ми хочемо дізнатися, які втрати може спричинити поява важливих змін у законодавстві. Для цього потрібно ввести у мережу доказ (свідчення) `Changes_in_legislation = "Affect"`. Після виконання пропагації отримуємо наступні апостеріорні розподіли (Рисунок 15).

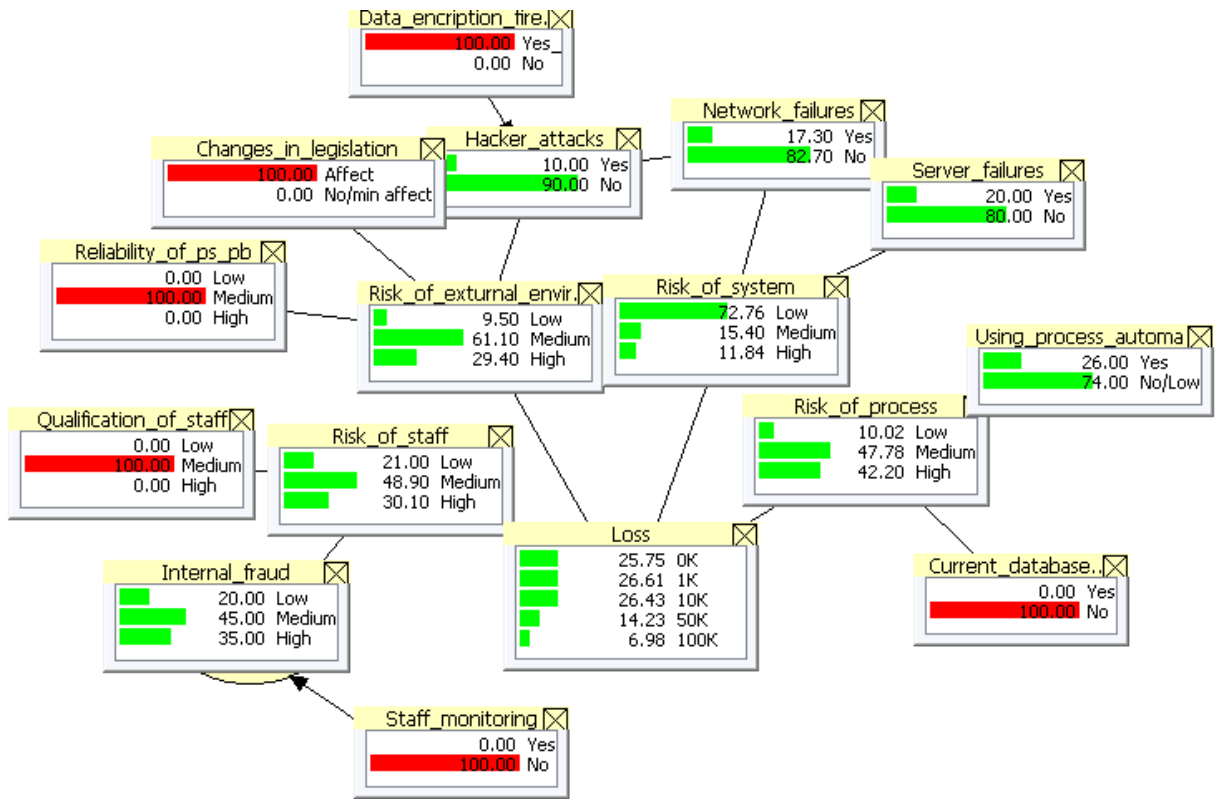


Рисунок 15 – Апостеріорні імовірності сценарного аналізу

У цьому випадку необхідний капітал для покриття операційного ризику становить 64183 долари.

Таким чином, побудовано мережу, що відображає процеси, притаманні даному комерційному банку, дає змогу оцінити капітал, необхідний для покриття ризику, та проводити сценарний аналіз.

### 3.5 Оцінювання та прогнозування операційних ризиків за допомогою дерев рішень

Як альтернативний метод моделювання та оцінювання операційних ризиків спробуємо застосувати дерева рішень, оскільки дані методи є популярними та широко використовуються. Як і у випадку використання мережі Байєса, класифікуватимемо втрати за п'ятьма категоріями.

#### 3.5.1 Підготовка даних та вибір метрик

Оскільки дані, які б відображали причинно-наслідкові зв'язки між змінними, визначеними як фактори впливу на розміри капіталу на покриття операційних ризиків для нашого комерційного банку, відсутні, спробуємо вирішити цю проблему.

Для побудови дерева рішень використаємо дані, отримані за допомогою створеної раніше мережі Байєса. Виконавши `sampling`, отримали датасет із 10000 рядків.

Як було зазначено раніше, дерева рішень можуть використовуватися як для категоріальних, так і неперервних даних, проте методи, запропоновані у бібліотеках Python, не підтримують роботу із нечисловими даними. Тому потрібно перетворити датасет для подальшого використання.

Маємо два способи це зробити:

- а) позначити значення змінних певними числами;
- б) використати one-hot-encoding.

Застосуємо перший спосіб. Отож, маємо наступні значення змінних (Таблиця 5).

Таблиця 5 – Значення змінних

№	Значення до перетворення	Присвоєне значення
1	Yes	1
2	No	0
3	High	2
4	Medium	1
5	Low	0

Задачу класифікації потрібно вирішити для втрат (Loss), які і є цільовою змінною. Цільова змінна має п'ять класів, які перетворимо також (Таблиця 6).

Таблиця 6 – Значення цільової змінної

№	Значення до перетворення	Присвоєне значення
1	0K	0
2	1K	1
3	10K	2
4	50K	3
5	100K	4

Як бачимо, дана задача відноситься до задач класифікації із кількома класами (Multi-class classification problem), адже цільова змінна не є бінарною.

Перевіримо збалансованість наших даних (Рисунок 16).

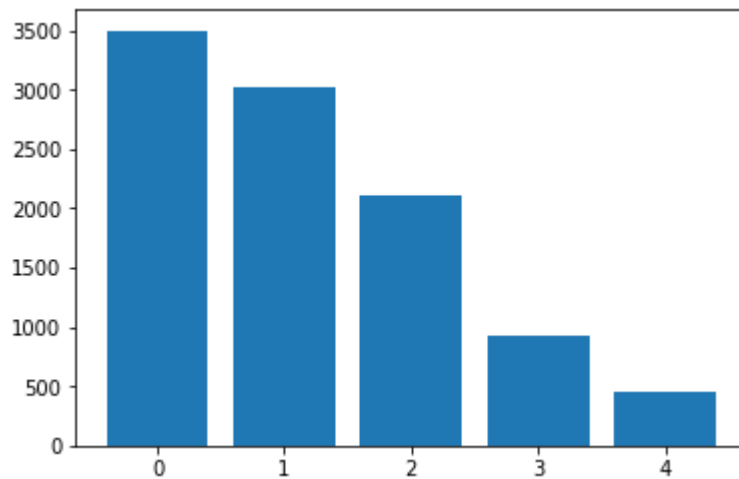


Рисунок 16 – Кількість записів для значень цільової змінної

Як бачимо, дані не збалансовані. Для збалансування даних догенеруємо записи для всіх значень цільової змінної, крім 0.

Отримаємо збалансовані дані (Рисунок 17).

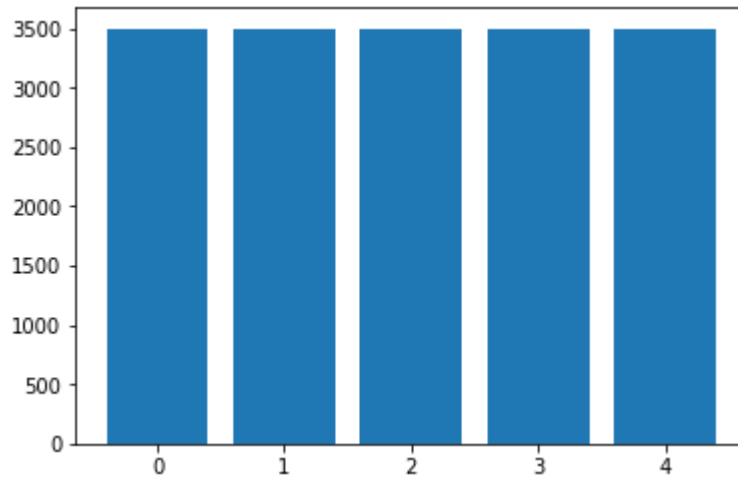


Рисунок 17 – Збалансовані дані

Для оцінки якості моделі необхідно обрати метрику, яку в подальшому максимізуватимемо. Оскільки дані збалансовано, то проблем у застосуванні точності (accuracy) немає. Тож для моніторингу оберемо точність та зважене усереднене F1-Score.

Метрика F1-Score є середнім гармонійним влучності та повноти. Влучність (Precision) визначається за формулою:

$$\text{Precision} = TP / (TP + FP),$$

де  $TP$  – елементи, справжнє значення яких є позитивним, і передбачене значення яких також позитивне,

$FP$  – елементи, справжнє значення яких є негативним, проте передбачене значення є позитивним.

Повнота (Recall) визначається за формулою:

$$\text{Recall} = TP / (TP + FN),$$

де  $FN$  – елементи, справжнє значення яких є позитивним, проте воно неправильно передбачене як негативне.

Таким чином, F1-Score визначається за формулою:

$$F1 - \text{Score} = (2 * \text{precision} * \text{recall}) / (\text{precision} + \text{recall}).$$

Метрики можуть бути усереднені трьома способами:

- макро-усереднення – незважене усереднене для кожного класу;
- зважене усереднене – усереднення з урахуванням ваг для кожного класу;
- мікро-усереднене – усереднення загальної кількості справжніх позитивних, помилкових негативних та помилкових позитивних результатів.

Для обчислення загальної точності потрібно додати елементи на діагоналі матриці невідповідностей і поділити на суму усіх елементів матриці.

Після підготовки даних варто оцінити нижню межу можливої точності моделі. Для цього використаємо наївний класифікатор Zero Rule Algorithm, за яким необхідно визначити значення цільової змінної, яке найчастіше зустрічається у даних. У даному випадку кількість всіх значень однакова, тому обираємо перше – 0. Це значення присвоюється всім цільовим змінним тестових записів, порівнявши які зі справжніми значеннями, отримаємо мінімальну точність моделі. У нашому випадку це 0,194.

### 3.5.2 Побудова дерева рішень та аналіз результатів

Після обробки даних, вибору метрик та оцінки нижньої межі точності прогнозу для наших даних можемо будувати дерево рішень. Для цього використаємо алгоритм Classification And Regression Tree (CART).

Отже, потрібно класифікувати втрати (Loss) до одного із п'яти класів: 0, 1K, 10K, 50K, 100K.

Розділення вузлів будемо виконувати за методом Gini Impurity.

Спочатку побудуємо дерево для наших даних, використавши параметри, встановлені за замовчуванням. Отримали наступні результати (Рисунок 18).



```

test: 0.6276078879679908

              precision    recall  f1-score   support

     0       0.547        0.558        0.552         735
     1       0.462        0.398        0.427         684
     2       0.516        0.516        0.516         674
     3       0.737        0.771        0.754         682
     4       0.829        0.884        0.856         724

 accuracy          0.628         3499
 macro avg       0.618        0.625        0.621         3499
 weighted avg    0.620        0.628        0.623         3499

```

Рисунок 18 – Результати застосування дерева рішень із параметрами за замовчуванням

Спробуємо покращити точність, підібравши параметри за допомогою `RandomizedSearchCV`, а потім уточнити їх, використавши `GridSearchCV`.

Зазвичай, ми маємо лише неясне уявлення про найкращі гіперпараметри, і, отже, найкращим підходом до звуження нашого пошуку є оцінка широкого діапазону значень для кожного гіперпараметра. Використовуючи метод `RandomizedSearchCV` `Scikit-Learn`, ми можемо визначити сітку діапазонів гіперпараметрів та випадково взяти вибірку з сітки, виконуючи K-Fold CV з кожною комбінацією значень.

Випадковий пошук дозволив нам звужити діапазон для кожного гіперпараметра. Тепер, коли ми знаємо, де зосередити свій пошук, ми можемо чітко вказати кожну комбінацію налаштувань для подальшого пошуку. Ми робимо це за допомогою `GridSearchCV`, методу, який замість випадкової вибірки з розподілу обчислює всі визначені нами комбінації. Для використання `Grid Search` ми робимо іншу сітку на основі найкращих значень, отриманих за випадковим пошуком [34].

Спробуємо побудувати випадковий ліс.

Випадковий ліс має кращу прогножуючу потужність та точність, ніж одна модель CART. На відміну від моделі CART, правила випадкового лісу нелегко інтерпретувати [35].

Отже, побудуємо дерево за допомогою випадкового лісу із параметрами, встановленими за замовчуванням, а також спробуємо підібрати гіперпараметри вище описаними методами.

Отримані результати представлено у таблиці 7.

Таблиця 7 – Результати застосування CART та Random Forest

Метод	max_ depth	max_ feature s	min_ samp les_ leaf	min_ samp les_ split	n_ est imat ors	bootst rap	criteri on	accuracy	F1 weighte d avg
CART default param.	None	None	1	2	-	-	gini	0.617	0.612
CART	22	15	1	2	-	-	gini	0.628	0.623
Rando m Forest def. par.	None	None	1	2	100	True	-	0.649614 1754787 083	0.642
Rando m Forest	60	'log2'	2	2	230	False	-	0.659617 0334381 252	0.651

Отже, для CART з підібраними параметрами маємо такі результати:

test: 0.6276078879679908

	precision	recall	f1-score	support
0	0.547	0.558	0.552	735
1	0.462	0.398	0.427	684
2	0.516	0.516	0.516	674

3	0.737	0.771	0.754	682
4	0.829	0.884	0.856	724
accuracy			0.628	3499
macro avg	0.618	0.625	0.621	3499
weighted avg	0.620	0.628	0.623	3499

Матриця невідповідностей має наступний вигляд (Рисунок 19):

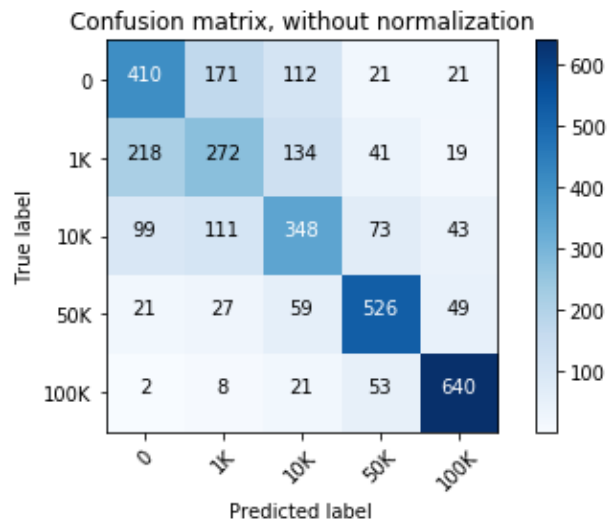


Рисунок 19 – Матриця невідповідностей для CART із підібраними гіперпараметрами

Для кількісної оцінки корисності всіх змінних у всьому випадковому лісі використаємо відносну важливість змінних. Значення у таблиці відображають, на скільки включення певної змінної покращує прогнозування (Рисунок 20) [36].

	FeatureImportance
Internal_fraud	0.104358
Qualification_of_staff	0.101235
Risk_of_process	0.089358
Risk_of_system	0.085864
Risk_of_staff	0.084791
Reliability_of_ps_pb	0.078458
Changes_in_legislation	0.072541
Staff_monitoring	0.072326
Risk_of_exturnal_environment	0.067342
Current_database_of_business_processes	0.051467
Using_process_automation	0.044107
Data_encryption_firewall	0.038427
Server_failures	0.037202
Hacker_attacks	0.036442
Network_failures	0.036084

Рисунок 20 – Важливість змінних

Як бачимо, отримана точність моделі не є дуже високою, проте перевищує нижній поріг.

Спробуємо покращити точність, застосувавши випадковий ліс. Після підбору параметрів отримали такі результати:

```
test: 0.6596170334381252
```

	precision	recall	f1-score	support
0	0.605	0.595	0.600	735
1	0.521	0.414	0.461	684
2	0.559	0.537	0.548	674
3	0.726	0.824	0.772	682
4	0.818	0.917	0.865	724
accuracy			0.660	3499
macro avg	0.646	0.657	0.649	3499
weighted avg	0.647	0.660	0.651	3499

Побудуємо матрицю невідповідностей (Рисунок 21).

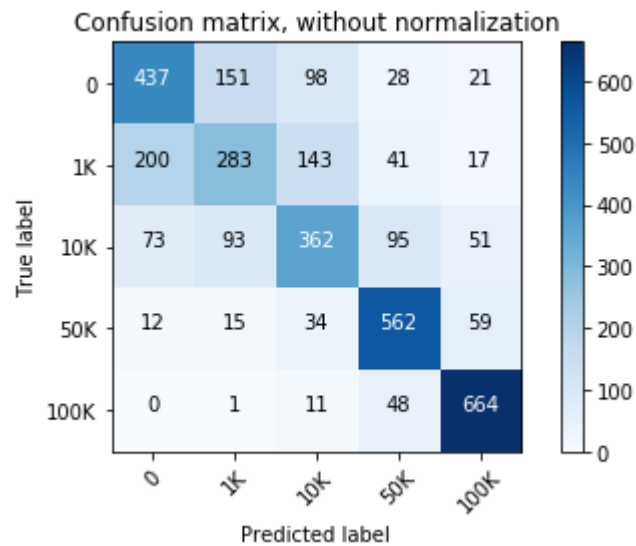


Рисунок 21 – Матриця невідповідностей для Random Forest із підібраними гіперпараметрами

Маємо наступну важливість змінних (Рисунок 22):

Risk_of_staff	Importance: 0.13
Risk_of_exturnal_environment	Importance: 0.12
Risk_of_process	Importance: 0.12
Risk_of_system	Importance: 0.1
Internal_fraud	Importance: 0.09
Qualification_of_staff	Importance: 0.09
Reliability_of_ps_pb	Importance: 0.07
Staff_monitoring	Importance: 0.05
Changes_in_legislation	Importance: 0.05
Current_database_of_business_processes	Importance: 0.04
Data_encryption_firewall	Importance: 0.03
Network_failures	Importance: 0.03
Server_failures	Importance: 0.03
Using_process_automation	Importance: 0.03
Hacker_attacks	Importance: 0.02

Рисунок 22 – Важливість змінних

Отже, алгоритми CART та випадковий ліс було застосовано до згенерованих на основі побудованої раніше мережі Байєса даних. Оскільки мережа будувалася за допомогою експертних даних, то і згенерований датасет містив лише

категоріальні дані, що не зовсім відповідає реальній задачі, адже на практиці для обчислення капіталу на покриття операційного ризику переважно використовуються числові дані та розподіли втрат. Проте за відсутності необхідних даних було використано наявний датасет з перетворенням записів на числові. Оскільки значенням змінних, зокрема і цільової, було присвоєно так званий код, який не несе числової інформації, це могло вплинути на точність алгоритмів. Адже з наведених матриць невідповідностей бачимо, що найкраще методи класифікують записи із значенням цільової змінної 100K, що має найбільше значення числа-коду. У цьому випадку можна застосувати one-hot encoding, що, проте, збільшить об'єм даних. Було проведено підбір гіперпараметрів моделей, що трохи покращило результати. Для більшої ефективності потрібні більш репрезентативні дані. Також можна провести дослідження із застосуванням інших методів.

### 3.6 Висновки до розділу 3

У даному розділі наведено обґрунтування вибору платформи для побудови мережі Байєса та реалізації дерев рішень, вказано вимоги до обладнання та інструменти для роботи з даними. Також наведено етапи підготовки даних, побудови мережі та дерев рішень, представлено результати моделювання та оцінювання операційних ризиків за допомогою мережі Байєса та дерев рішень. Дані моделі дають змогу оцінити необхідний капітал на покриття операційних ризиків у комерційному банку.

## РОЗДІЛ 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

У цьому розділі проводиться розробка стартап-проекту на основі створеної мережі Байєса для оцінювання капіталу, необхідного на покриття ОР.

### 4.1 Опис ідеї стартап-проекту

Початком стартап-проекту є ідея, яку потрібно чітко сформулювати. Опис ідеї даного стартап-проекту представлений у таблиці 8.

Таблиця 8– Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Система для оцінювання розміру необхідного капіталу на покриття ОР комерційного банку на основі БМ.	Оцінювання розміру необхідного капіталу на покриття ОР комерційного банку.	<p>Дає змогу швидко оцінити розмір необхідного капіталу;</p> <p>зручна і наглядна у користуванні навіть без значного досвіду у роботі з БМ;</p> <p>дозволяє проводити сценарний аналіз і визначати можливі втрати за різних ситуацій;</p> <p>отримані результати можуть слугувати основою для подальшого прийняття рішень.</p>

Проаналізуємо дану ідею для прийняття рішення про доцільність її подальшої реалізації. Аналіз ідеї проведено у таблиці 9.

Таблиця 9 – Сильні, слабкі та нейтральні характеристики ідеї стартап-проекту

№ п/п	Техніко-економічні характеристики ідеї	(Потенційні) товари/концепції конкурентів			
		Моя система	LogicGate Risk Cloud	Fusion Framework System	SpheraCloud
1	Ціна	Низька	Висока	Безкоштовна	Висока
2	Функціонал	Вузький	Широкий	Широкий	Надширокий

З таблиці 9 можемо бачити, що, загалом, ціна є сильною характеристикою системи. Зважаючи на область застосувань і основне завдання системи, функціонал також може бути сильною характеристикою, оскільки конкурентна безкоштовна система не є простою у використанні і потребує кваліфікованих працівників для роботи із нею.

#### 4.2 Розробка бізнес-моделі стартапу

Наступним кроком є розробка бізнес-моделі проекту. Вона наведена у таблиці 10.



Таблиця 10 - Бізнес-модель Canvas

№ п/п	Назва	Зміст
1	Споживчі сегменти	Нішовий ринок (комерційні банки).
2	Ключові види діяльності	Система для оцінювання необхідного капіталу на покриття ОР.
3	Ціннісна пропозиція	Зручність – інтуїтивно зрозуміле, зручне та швидке представлення результатів та введення даних. Доступність: має низьку ціну, малий обсяг необхідних комп'ютерних ресурсів. Може використовуватися працівниками із будь-яким рівнем підготовки. Дає змогу розглядати різні сценарії. Може бути основою для прийняття подальших управлінських рішень.
4	Канали збуту	Демоверсія системи, реклама.
5	Взаємовідносини з клієнтами	Персональна підтримка.
6	Потоки надходження доходу	Ліцензія.
7	Ключові ресурси	Матеріальні (комп'ютери, інтернет, програмне середовище для створення БМ). Інтелектуальні (моделі, алгоритми, дані). Людські (бізнес аналітик, розробники, тестувальник, експерт - консультант). Фінансові (гроші на розвиток та підтримку системи, ліцензія на комерційне використання програмного середовища для створення БМ).

Кінець таблиці 10

№ п/п	Назва	Зміст
8	Ключові партнери	Відносини виробника з постачальниками (розробники програмного забезпечення для побудови БМ) для гарантії отримання якісних комплектуючих.
9	Структура витрат	З переважною увагою до цінності. Витрати: фіксовані (ліцензія на використання ПЗ). Змінні (тех. підтримка, збереження даних, реклама).

#### 4.3 Аналіз ринкових можливостей та розробка маркетингової стратегії стартап-проекту

У ході аналізу ринкових можливостей розглянемо характеристики потенційних клієнтів нашого стартап-проекту (таблиця 11), визначимо ринкові можливості та загрози (таблиця 12), проведемо аналіз конкуренції в обраній галузі за М. Портером (таблиця 13) та здійснимо обґрунтування факторів конкурентоспроможності (таблиця 14). Підсумуємо все це у SWOT-аналізі стартап-проекту (таблиця 15).

Таблиця 11 - Характеристика потенційних клієнтів стартап-проекту

Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
Необхідна оцінка капіталу на покриття ОР у різних ситуаціях для прийняття управлінських рішень.	Комерційні банки.	Ціновий фактор, функціонал, наявність ліцензії, умови договору з банком.	Якісні послуги та надійні моделі. Зрозумілість, зручність використання системи. Релевантна ціна.

Таблиця 12 - Визначення ринкових можливостей і загроз

Параметри оцінки	Можливості	Загрози
1. Конкуренція	Малі банки не можуть собі дозволити дороге програмне забезпечення, не мають кваліфікованих спеціалістів для роботи з ним.	Поява безкоштовних платформ для оцінювання операційних втрат.
2. Збут	Висока комп'ютеризованість.	-

Кінець таблиці 12

Параметри оцінки	Можливості	Загрози
3.Політичні і правові чинники	- .	Введення використання спеціального ПЗ для всіх банків України.
4.Міжнародні чинники	Необхідність точної оцінки та розгляду сценаріїв у зв'язку з пандемією.	Всесвітня криза.

Таблиця 13 - Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
Складові аналізу	SpheraCloud	Бар'єри входження в ринок: затрати, ресурси, попит	Підписання контрактів в про співпрацю	Попит	Функціонал
Висновки:	Низька інтенсивність конкурентної боротьби з боку прямих конкурентів.	-є можливості входу в ринок; - немає потенційних конкурентів.	Постачальники не диктують умови роботи на ринку.	Клієнти диктують такі умови роботи на ринку: якість (точність результатів), зручність у використанні, доступність.	Обмеження для роботи на ринку через товари заміни: обмеження у збалансованості ціни та функціоналу.

Таблиця 14 - Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування
1	Вузька спеціалізація	Не потрібно платити за функції системи, які не використовуються.
2	Простота у використанні	Продукт максимально user friendly та не перевантажений зайвою інформацією.
3	Вартість	Не завищена, конкурентна ціна.
4	Рівень якості послуг	Система дає якісну оцінку необхідного капіталу.

Таблиця 15 - SWOT- аналіз стартап-проекту

	<b>Сильні сторони</b> <b>1. Висока якість системи.</b> <b>2. Зручність та простота у використанні.</b> <b>3. Релевантна ціна.</b>	<b>Слабкі сторони</b> <b>1.Відсутність репутації.</b> <b>2.Неможливість використання без комп'ютера з встановленим необхідним ПЗ.</b> <b>3.Вузький функціонал.</b>
<b>Можливості</b> <b>1.Поява попиту на системи підтримки прийняття рішень у зв'язку з пандемією.</b>	З появою попиту на СППР дана система, завдяки релевантній ціні, високій якості послуг та зручності у використанні займе вигідну позицію на ринку.	Відсутність репутації можна компенсувати за рахунок появи попиту через пандемію.

Кінець таблиці 15.

<b>2. Висока комп'ютеризація банків.</b>	Завдяки високій комп'ютеризованості банків збільшиться кількість потенційних клієнтів.	Завдяки високій комп'ютеризації банків проблем необхідності комп'ютера для користування системою не буде серйозною.
<b>Загрози</b> <b>1. Вхід на український ринок закордонних постачальників.</b> <b>2. Обмеженість ресурсів малих банків внаслідок пандемії.</b>	Помірна ціна буде сильною стороною в порівнянні з іноземними системами. Висока якість системи та релевантна ціна привернуть увагу та зроблять доступною систему у період пандемії.	Збільшення довіри та попиту допоможе швидко набути репутацію. Вузький функціонал обумовлює нижчу ціну.

#### 4.4 Розробка маркетингової програми стартап-проекту

Етапи розробки маркетингової програми стартап-проекту наведено у таблицях 16 – 21.

Таблиця 16 - Визначення базової стратегії розвитку

Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
Цільовий маркетинг	Якісна, зручна, проста у використанні система з неперевантаженим функціоналом за релевантною ціною.	Позиціонування за співвідношенням «ціна-якість»

Таблиця 17 - Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
Ні	Шукати нових та частково забирати існуючих	Не буде	Стратегія заняття конкурентної ніші

Таблиця 18 - Визначення ключових переваг концепції потенційного товару

Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
Необхідна оцінка капіталу на покриття ОР у різних ситуаціях для прийняття управлінських рішень.	Швидка, надійна, зрозуміла та зручна у використанні система.	Низька ціна.

Таблиця 19 - Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові
I. Товар за задумом	Швидкий аналіз даних та визначення оцінки розміру необхідного капіталу.
II. Товар у реальному виконанні	Система, яка за введеними даними визначає ймовірність потрапляння розміру втрат у певний проміжок.
	Малий розмір ПЗ.
	Інтуїтивно зрозумілий інтерфейс, зручне введення інформації.
III. Товар із підкріпленням	До продажу: наявність комп'ютера зі встановленим необхідним ПЗ.
	Після продажу: ліцензія на використання, технічна підтримка.



Таблиця 20 - Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
Клієнти не користуються системами підтримки прийняття рішень або купують дороге ПЗ, часто не використовуючи весь доступний функціонал.	Встановлення контактів зі споживачами та їх підтримка.	Канал нульового рівня.	Реклама системи, демоверсія.

Таблиця 21 - Концепція маркетингових комунікацій

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
Клієнти дізнаються про систему через рекламу, на спеціальних з'їздах та форумах.	Інтернет, спільноти працівників банків.	Контент-маркетинг, публікації у спеціалізованій літературі.	Інформувати про існування, можливості та переваги даної системи.	Надійна оцінка без переплат за непотрібні функції.

#### 4.5 Висновки до розділу 4

У цьому розділі проаналізовано можливість впровадження створеної системи у вигляді стартап-проекту. Для цього здійснено аналіз ідеї, розроблено бізнес-модель, проаналізовано ринкові можливості та створено маркетингову програму стартапу.

У результаті проведеного аналізу можемо сказати, що, системи для роботи із операційними ризиками існують, проте вони мають високу ціну або перевантажені функціоналом, що означає додаткові витрати на навчання працівників. Враховуючи ситуацію у світі, оцінювання операційних ризиків буде одним із основних завдань банків. Тому наша система має шанси привернути увагу достатньо широкого кола споживачів. Програмні характеристики створеного продукту, інтерфейс та ціна будуть орієнтуватися на українського споживача. Таким чином, навіть при входженні на ринок іноземних постачальників, продукт залишиться конкурентоспроможним. Таким чином подальша імплементація проекту є доцільною.

## ВИСНОВКИ

У ході написання магістерської дисертації виконано дослідження актуальності проблеми моделювання та оцінювання операційних ризиків. Враховуючи нестабільність в усьому світі, спричинену поширенням COVID-19 та введенням запобіжних заходів, що вже призвело до значних втрат та банкрутства багатьох установ, важливо приймати правильні управлінські рішення на основі якісних моделей та оцінок. Також у зв'язку з поширенням COVID-19 на роботу установ посилився вплив проблем інформаційної безпеки, ведення документації та шахрайства, пов'язаного з кіберпростором. Тож моделювання і оцінювання операційних ризиків є актуальною проблемою.

Аналіз літератури та існуючих результатів показав, що опубліковано багато досліджень застосування методів, описаних у Базелі II, а також операційних ризиків в цілому. Також досить популярним є моделювання операційних ризиків за допомогою різних видів мереж Байєса. Застосування ж дерев рішень не повністю досліджене, проте є перспективним.

Для побудови мережі Байєса для оцінки ОР комерційного банку було використано експертні судження. Мережа відображає процеси, притаманні даному комерційному банку, дає змогу розглядати різні сценарії, відтворити поточну ситуацію за допомогою введення доказів та на основі цих даних отримати оцінку необхідного капіталу на покриття ОР. Мережу було створено за допомогою програмного забезпечення Hugin Lite.

Дерева рішень CART, випадковий ліс було навчено на згенерованих за побудованою мережею даних, оскільки у відкритому доступі реальних даних, які б коректно відображали операційні втрати комерційного банку, не було знайдено.

За результатами оцінювання ОР за мережею Байєса можемо зробити висновок, що дана мережа може використовуватися у комерційних банках.

Результати застосування дерев рішень та випадкового лісу не є достатніми для впровадження цих моделей у роботу банків і потребують покращення. Основною проблемою, що спричинила низькі показники обраних метрик, вважаємо не достатньо якісні і репрезентативні дані.

В подальшій роботі необхідно звернути увагу на такі питання:

- підбір фактичних або більш репрезентативних даних для навчання мережі та дерев рішень та порівняння з відомими розробленими моделями;
- покращення розробленої мережі Байєса за допомогою поглиблення власних знань предметної області та консультації експертів, комбінування експертних суджень з реальними даними;
- використання one-hot encoding та дослідження інших можливих методів покращення роботи дерев рішень;
- дослідження застосування інших методів машинного навчання для оцінювання операційних ризиків;
- побудова спеціалізованої системи підтримки прийняття рішень для аналізу операційних та інших фінансових ризиків.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Балабанов И.Т. Риск-менеджмент. Москва: Финансы и статистика, 1996. 192 с.
2. Фінансовий ризик. URL: [https://uk.wikipedia.org/wiki/%D0%A4%D1%96%D0%BD%D0%B0%D0%BD%D1%81%D0%BE%D0%B2%D0%B8%D0%B9\\_%D1%80%D0%B8%D0%B7%D0%B8%D0%BA](https://uk.wikipedia.org/wiki/%D0%A4%D1%96%D0%BD%D0%B0%D0%BD%D1%81%D0%BE%D0%B2%D0%B8%D0%B9_%D1%80%D0%B8%D0%B7%D0%B8%D0%BA) (accessed: 10.09.2020).
3. Операційний ризик. URL: [https://uk.wikipedia.org/wiki/%D0%9E%D0%BF%D0%B5%D1%80%D0%B0%D1%86%D1%96%D0%B9%D0%BD%D0%B8%D0%B9\\_%D1%80%D0%B8%D0%B7%D0%B8%D0%BA](https://uk.wikipedia.org/wiki/%D0%9E%D0%BF%D0%B5%D1%80%D0%B0%D1%86%D1%96%D0%B9%D0%BD%D0%B8%D0%B9_%D1%80%D0%B8%D0%B7%D0%B8%D0%BA) (accessed: 10.09.2020).
4. Operational risk. URL: [https://en.wikipedia.org/wiki/Operational\\_risk](https://en.wikipedia.org/wiki/Operational_risk) (accessed: 10.09.2020).
5. Дмитрова О.С., Гончарова К.Г., Меренкова О.В., Медвідь Т.А., Бойко А.О., Вахнюк С.В. Моделювання оцінки операційного ризику комерційного банку: монографія / за заг. ред. С.О. Дмитрова. Суми: Державний вищий навчальний заклад «Українська академія банківської справи Національного банку України», 2010. 277 с.
6. Операционный риск. URL: [https://ru.wikipedia.org/wiki/%D0%9E%D0%BF%D0%B5%D1%80%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B9\\_%D1%80%D0%B8%D1%81%D0%BA](https://ru.wikipedia.org/wiki/%D0%9E%D0%BF%D0%B5%D1%80%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B9_%D1%80%D0%B8%D1%81%D0%BA) (accessed: 10.09.2020).
7. COVID-19: Op risk forum identifies top-three controls weakened by home working, notes shift to crisis management. URL: <https://blogs.thomsonreuters.com/answerson/covid-19-op-risk-forum->

- identifies-top-three-controls-weakened-by-home-working-notes-shift-to-crisis-management/ (accessed: 12.09.2020).
8. Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework. URL: <https://www.bis.org/publ/bcbs107.htm> (accessed: 12.09.2020).
  9. Reason James. Managing the Risks of Organizational Accidents. Aldershot: Ashgate Publishing Limited, 1997. 252 p.
  10. David Häger , Martin Neil, Lasse B. Andersen. Modelling Operational Risk in Financial Institutions using Hybrid Dynamic Bayesian Networks. URL: [https://www.researchgate.net/publication/228356143\\_Modelling\\_Operational\\_Risk\\_in\\_Financial\\_Institutions\\_using\\_Hybrid\\_Dynamic\\_Bayesian\\_Network](https://www.researchgate.net/publication/228356143_Modelling_Operational_Risk_in_Financial_Institutions_using_Hybrid_Dynamic_Bayesian_Network).
  11. Jöhenmark A. Modeling Operational Risk: Master of Science Thesis / Royal Institute of Technology. Sweden, Stockholm, 2012. 68 p.
  12. Alexander Carol. Bayesian Methods for Measuring Operational Risk. URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=248148](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=248148).
  13. Alexander Carol. Operational risk, Regulation, Analysis and Management. Midsomer Norton: Bookcraft Ltd, 2003. 369 p.
  14. Karam E., Planchet F. Operational Risks in Financial Sectors. *Advances in Decision Sciences*. URL: <https://www.hindawi.com/journals/ads/2012/385387/>.
  15. Edoh Fofo Afambo. Operational Risk Capital Provisions for Banks and Insurance Companies: Doctoral dissertation / Georgia State University. USA, Atlanta, 2006. 165 p.
  16. Best Operational Risk Management Software. URL: [https://www.g2.com/categories/operational-risk-management?utf8=%E2%9C%93&order=top\\_shelf](https://www.g2.com/categories/operational-risk-management?utf8=%E2%9C%93&order=top_shelf) (accessed: 21.09.2020).
  17. LogicGate Risk Cloud. URL: <https://www.g2.com/products/logicgate-risk-cloud/reviews> (accessed: 21.09.2020).

18. Fusion Framework System. URL: <https://www.g2.com/products/fusion-framework-system/reviews> (accessed: 21.09.2020).
19. SpheraCloud. URL: <https://sphera.com/spheracloud/> (accessed: 21.09.2020).
20. Панкратова Н.Д., Бідюк П.І., Рубець М.Г. Інформаційна система для моделювання та оцінювання фінансових операційних ризиків за допомогою байєсівської мережі. *Системні дослідження та інформаційні технології*. 2015. № 3. С. 7 – 19.
21. Згуровський М. З., Бідюк П. І., Терентьев О. М., Просянкін-Жарова Т. І. Байєсівські мережі в системах підтримки прийняття рішень. Київ: ТОВ «Видавниче Підприємство «Едельвейс», 2015. 300 с.
22. Бідюк П.І., Терентьев О.М., Коновалюк М.М. Байєсівські мережі в технологіях інтелектуального аналізу даних. *Штучний інтелект*. 2010. № 2. С. 104 – 113.
23. Greedy Search-And-Score Structure Learning. URL: [http://data.biotracer.hugin.com/htmlhelp/descr\\_greedy\\_pane.html](http://data.biotracer.hugin.com/htmlhelp/descr_greedy_pane.html) (accessed: 29.09.2020).
24. Padmanaban Harini. Comparative Analysis of Naive Bayes and Tree Augmented Naive Bayes Models: Master's Theses / San Jose State University. USA, San Jose, 2014. 65 p.
25. Chow, C.K. C.N. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Trans. on Info. Theory*. 1968. №14. P. 462- 467.
26. Kwabena Adusei-Poku. Operational Risk Management – Implementing a Bayesian Network for Foreign Exchange and Money Market Settlement: Doctoral dissertation / University of Gottingen. Germany, Gottingen, 2005. 152 p.
27. Yew Khuen Yoon. Modelling Operational Risk in Financial Institutions Using Bayesian Networks: Master's Theses / Cass Business School. Great Britain, London, 2003. 83 p.

28. Martin Leo, Suneel Sharma, K. Maddulety. Machine Learning in Banking Risk Management: A Literature Review. *Risks*. Sydney, 2019. No. 7, 29. P. 1-22.
29. Decision tree learning. URL: [https://en.wikipedia.org/wiki/Decision\\_tree\\_learning](https://en.wikipedia.org/wiki/Decision_tree_learning).
30. 4 Simple Ways to Split a Decision Tree in Machine Learning. URL: <https://www.analyticsvidhya.com/blog/2020/06/4-ways-split-decision-tree/>.
31. Random Forest. URL: <https://corporatefinanceinstitute.com/resources/knowledge/other/random-forest/>.
32. Random Forest. URL: [https://uk.wikipedia.org/wiki/Random\\_forest](https://uk.wikipedia.org/wiki/Random_forest).
33. Pearl J., Russel S. Bayesian networks / ed. M. Arbib. *In Handbook of Brain Theory and Neural Networks: Report (R-277)*, November 2000. MIT Press, Cambridge, 2001. P. 157–160.
34. Hyperparameter Tuning the Random Forest in Python. URL: <https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74>.
35. CART and Random Forests for Practitioners. URL: <https://dzone.com/articles/cart-and-random-forests>.
36. Random Forest in Python. URL: <https://towardsdatascience.com/random-forest-in-python-24d0893d51c0>.
37. Classification And Regression Trees for Machine Learning. URL: <https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/>.



## ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

```

from pandas import read_csv
from imblearn.over_sampling import SMOTE
from collections import Counter
from sklearn.preprocessing import LabelEncoder
import pandas as pd
import numpy as np
from scipy.sparse import csr_matrix
from sklearn import tree
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics

data = pd.read_csv(r"d:\disk1\...\or_cases.csv")
data_output = data['Loss']
d2 = data.copy()
del d2['Loss']
data_input = d2.copy()

data_output = LabelEncoder().fit_transform(data_output)
oversample = SMOTE()
data_input, data_output = oversample.fit_resample(data_input, data_output)

X_train, X_test, Y_train, y_test = train_test_split(data_input, data_output, test_size=0.2)

from sklearn.model_selection import RandomizedSearchCV
splitter = ['best', 'random']
max_features = ['auto', 'sqrt']
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]

```

```

max_depth.append(None)
min_samples_split = [2, 5, 10]
min_samples_leaf = [1, 2, 4]
criterion = ['gini', 'entropy']
random_grid = {'splitter': splitter,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'criterion': criterion}

rf = DecisionTreeClassifier()
rf_random = RandomizedSearchCV(estimator = rf, scoring = 'f1_micro', param_distributions =
random_grid, n_iter = 100, cv = 3, verbose=2, random_state=100, n_jobs = -1)
rf_random.fit(X_train, Y_train)
rf_random.best_params_

from sklearn.model_selection import GridSearchCV
param_grid = {
    'splitter': ['best', 'random'],
    'max_depth': [13, 20, 30, 35, None],
    'max_features': ['sqrt', 4, 9, 10, 11, None],
    'min_samples_leaf': [1, 3],
    'min_samples_split': [2, 3, 5],
    'criterion': ['gini', 'entropy']
}
rf = DecisionTreeClassifier(random_state=100)
grid_search = GridSearchCV(estimator = rf, scoring = 'f1_micro', param_grid = param_grid,cv = 3,
n_jobs = -1, verbose = 2)
grid_search.fit(X_train, Y_train)
grid_search.best_params_

model = DecisionTreeClassifier(criterion = 'gini', max_features = 15, max_depth=None, splitter = 'best',
random_state = 100, min_samples_leaf=1, min_samples_split=2).fit(X_train, Y_train)
y_pred = 0

```

```

y_pred = model.predict(X_test)

print('test:', accuracy_score(y_test, y_pred))
cm = confusion_matrix(y_test, y_pred)
print(metrics.classification_report(y_test, y_pred, digits=3))

from sklearn.model_selection import RandomizedSearchCV
n_estimators = [int(x) for x in np.linspace(start = 10, stop = 1000, num = 10)]
max_features = ['log2', 'sqrt', None]
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
min_samples_split = [2, 5, 10]
min_samples_leaf = [1, 2, 4]
bootstrap = [True, False]
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}

rf = RandomForestClassifier()
rf_random = RandomizedSearchCV(estimator = rf, scoring = 'f1_micro', param_distributions =
random_grid, n_iter = 100, cv = 3, verbose=2, random_state=42, n_jobs = -1)
rf_random.fit(X_train, Y_train)

model = RandomForestClassifier(bootstrap = False, n_estimators = 98, random_state = 100, max_depth
= 30, min_samples_split = 10, min_samples_leaf = 1, max_features = 'sqrt').fit(X_train, Y_train)
y_pred = 0
y_pred = model.predict(X_test)

param_grid = {
    'bootstrap': [True, False],
    'max_depth': [55, 60, 65, None],
    'max_features': [3, 4, 'sqrt', 7],

```

```

        'min_samples_leaf': [2, 4],
        'min_samples_split': [2, 4],
        'n_estimators': [100, 200, 230]
    }
    rf = RandomForestClassifier()
    grid_search = GridSearchCV(estimator = rf, scoring = 'f1_micro', param_grid = param_grid, cv = 3,
n_jobs = -1, verbose = 2)
    grid_search.fit(X_train, Y_train)
    grid_search.best_params_

    model = RandomForestClassifier(bootstrap = False, n_estimators = 230, random_state = 100, max_depth
= 60, min_samples_split = 2, min_samples_leaf = 2, max_features = 3).fit(X_train, Y_train)
    y_pred = 0
    y_pred = model.predict(X_test)

    print('test:', accuracy_score(y_test, y_pred))
    cm = confusion_matrix(y_test, y_pred)
    print(metrics.classification_report(y_test, y_pred, digits=3))

```

## ДОДАТОК Б ТАБЛИЦІ АПРІОРНИХ ІМОВІРНОТЕЙ

Таблиця Б.1 – Розподіл для Staff\_monitoring

Yes	0.6
No	0.4

Таблиця Б.2 – Розподіл для Internal\_fraud

Staff_monitorig	Yes	No
Low	0.5	0.2
Medium	0.3	0.45
High	0.2	0.35

Таблиця Б.3 – Розподіл для Qualification\_of\_staff

Low	0.3
Medium	0.5
High	0.2

Таблиця Б.4 – Розподіл для Data\_encryption\_firewall

Yes_	0.87
No	0.13

Таблиця Б.5 – Розподіл для Reliability\_of\_ps\_pb

Low	0.33
Medium	0.61
High	0.06

Таблиця Б.6 – Розподіл для Changes\_in\_legislation

Affect	0.5
No/min affect	0.5

Таблиця Б.7 – Розподіл для Hacker\_attacks

Data_encryption	Yes_	No
Yes	0.1	0.94
No	0.9	0.06

Таблиця Б.8 – Розподіл для Network\_failures

Hacker_attack	Yes	No
Yes	0.83	0.1
No	0.17	0.9

Таблиця Б.9 – Розподіл для Server\_failures

Yes	0.2
No	0.8

Таблиця Б.10 – Розподіл для Using\_process\_automation

Yes	0.26
No/Low	0.74

Таблиця Б.11 – Розподіл для Current\_database\_of\_business\_processes

Yes	0.68
No	0.32

Таблиця Б.12 – Розподіл для Risk\_of\_staff

Qualification_o	Low			Medium			High		
Internal_audit	Low	Medium	High	Low	Medium	High	Low	Medium	High
Low	0.41	0.23	0.01	0.47	0.18	0.1	0.75	0.44	0.1
Medium	0.5	0.47	0.5	0.4	0.52	0.5	0.25	0.36	0.4
High	0.09	0.3	0.49	0.13	0.3	0.4	0	0.2	0.5

Таблиця Б.13 – Розподіл для Risk\_of\_system

Server_failures	Yes		No	
Network_failur	Yes	No	Yes	No
Low	0.15	0.2	0.2	1
Medium	0.5	0.45	0.45	0
High	0.35	0.35	0.35	0

Таблиця Б.14 – Розподіл для Risk\_of\_process

Current_datab	Yes		No	
Using_process	Yes	No/Low	Yes	No/Low
Low	0.83	0.3	0.3	0.03
Medium	0.17	0.5	0.5	0.47
High	0	0.2	0.2	0.5

Таблиця Б.15 – Розподіл для Risk\_of\_exturnal\_environment

Hacker_attack	Yes					
Changes_in_le	Affect			No/min affect		
Reliability_of_	Low	Medium	High	Low	Medium	High
Low	0	0.05	0.08	0.15	0.27	0.57
Medium	0.2	0.35	0.42	0.3	0.5	0.35
High	0.8	0.6	0.5	0.55	0.23	0.08

Кінець таблиці Б.15

Hacker_attack	No					
Changes_in_le	Affect			No/min affect		
Reliability_of_	Low	Medium	High	Low	Medium	High
Low	0.05	0.1	0.4	0.25	0.6	0.98
Medium	0.57	0.64	0.5	0.6	0.35	0.02
High	0.38	0.26	0.1	0.15	0.05	0







Продовження таблиці Б.16

Risk_of_st aff	Medium								
Risk_of_p roces	Medium								
Risk_of_s yste	Low			Medium			High		
Risk_of_e xturn	Low	Mediu m	High	Low	Mediu m	High	Low	Mediu m	High
0K	0.62	0.55	0.2	0.55	0.35	0.16	0.2	0.16	0.05
1K	0.28	0.3	0.32	0.3	0.36	0.27	0.32	0.27	0.08
10K	0.1	0.14	0.445	0.14	0.26	0.29	0.445	0.29	0.16
50K	0	0.01	0.035	0.01	0.03	0.18	0.035	0.18	0.46
100K	0	0	0	0	0	0.1	0	0.1	0.25

Продовження таблиці Б.16

Risk_of_st aff	Medium								
Risk_of_p roces	High								
Risk_of_s yste	Low			Medium			High		
Risk_of_e xturn	Low	Mediu m	High	Low	Mediu m	High	Low	Mediu m	High
0K	0.24	0.2	0.09	0.2	0.16	0.05	0.09	0.05	0.01
1K	0.55	0.32	0.1	0.32	0.27	0.08	0.1	0.08	0.05
10K	0.2	0.445	0.32	0.445	0.29	0.16	0.32	0.16	0.15
50K	0.01	0.035	0.35	0.035	0.18	0.46	0.35	0.46	0.44
100K	0	0	0.14	0	0.1	0.25	0.14	0.25	0.35

Продовження таблиці Б.16

Risk_of_st aff	High								
Risk_of_p roces	Low								
Risk_of_s yste	Low			Medium			High		
Risk_of_e xturn	Low	Mediu m	High	Low	Mediu m	High	Low	Mediu m	High
0K	0.4	0.24	0.1	0.24	0.2	0.09	0.1	0.09	0.02
1K	0.5	0.55	0.11	0.55	0.32	0.1	0.11	0.1	0.07
10K	0.1	0.2	0.44	0.2	0.445	0.32	0.44	0.32	0.15
50K	0	0.01	0.23	0.01	0.035	0.35	0.23	0.35	0.44
100K	0	0	0.12	0	0	0.14	0.12	0.14	0.32

Продовження таблиці Б.16

Risk_of_st aff	High								
Risk_of_p roces	Medium								
Risk_of_s yste	Low			Medium			High		
Risk_of_e xturn	Low	Mediu m	High	Low	Mediu m	High	Low	Mediu m	High
0K	0.24	0.2	0.09	0.2	0.16	0.05	0.09	0.05	0.01
1K	0.55	0.32	0.1	0.32	0.27	0.08	0.1	0.08	0.05
10K	0.2	0.445	0.32	0.445	0.29	0.16	0.32	0.16	0.15
50K	0.01	0.035	0.35	0.035	0.18	0.46	0.35	0.46	0.44
100K	0	0	0.14	0	0.1	0.25	0.14	0.25	0.35

Кінець таблиці Б.16

Risk_of_st aff	High								
Risk_of_p roces	High								
Risk_of_s yste	Low			Medium			High		
Risk_of_e xturn	Low	Mediu m	High	Low	Mediu m	High	Low	Mediu m	High
0K	0.1	0.09	0.02	0.09	0.05	0.01	0.02	0.01	0
1K	0.11	0.1	0.07	0.1	0.08	0.05	0.07	0.05	0.05
10K	0.44	0.32	0.15	0.32	0.16	0.15	0.15	0.15	0.05
50K	0.23	0.35	0.44	0.35	0.46	0.44	0.44	0.44	0.5
100K	0.12	0.14	0.32	0.14	0.25	0.35	0.32	0.35	0.4

## ДОДАТОК В НАУКОВІ ПУБЛІКАЦІЇ

1. Корнійчук А. С. Моделювання та оцінювання операційного фінансового ризику. Проблеми інформатизації: матеріали восьмої міжнародної науково-технічної конференції (Харків, 26 листопада – 27 листопада 2020 р.). Харків, 2020. Т. 3. С. 18.

2. Корнійчук А. С., Бідюк П. І. Моделювання та оцінювання операційного фінансового ризику. *Системні науки і кібернетика*. 2020. С. 106 - 122.